

Ventrem quoque ut habeat pilosum

Havaintoja Varron Res Rusticae -teoksen henkilöiden kielestä

Janne Niemi
Pro gradu -tutkielma
Latinan kieli ja Rooman kirjallisuus
Humanistinen tiedekunta
Helsingin yliopisto
Marraskuu 2020



Tiedekunta/Osasto – Fakultet/Sektion – Faculty Humanistinen tiedekunta		
Tekijä – Författare – Author Janne Niemi		
Työn nimi – Arbetets titel – Title Ventrem quoque ut habeat pilosum – Havaintoja Varron Res Rusticae -teoksen henkilöiden kielestä		
Oppiaine – Läroämne – Subject Latinan kieli ja Rooman kirjallisuus		
Työn laji – Arbetets art – Level Pro gradu	Aika – Datum – Month and marraskuu 2020	Sivumäärä– Sidoantal – Number of pages 61 s. + liitteet 13 s.
Tiivistelmä – Referat – Abstract <p>Tämän tutkielman aiheena on Marcus Terentius Varron (116-27 eaa) <i>Res Rusticae</i> -teoksen dialogimuotoisen tekstin henkilöhahmojen kieli. Keskeisenä tutkimuskysymyksenä on selvittää, onko teoksessa esiintyvien puhujien välillä havaittavissa kielellisiä eroja. Analyysin pohjana on teoksen kaikki kolme kirjaa kattava digitaalinen tekstikorpus. Aineisto on jäsennetty ja annotoitu puupankiksi, joka sisältää tiedot tekstin saneiden muoto-opillisista ominaisuuksista ja dependenssieliopin mukaisista syntaktisista suhteista. Annotoinnissa on hyödynnetty tietokoneavusteista, automaattista jäsennystä, jonka käyttökelpoisuutta käsillä olevan tutkimuskysymyksen analyysissä työssä myös arvioidaan.</p> <p><i>Res Rusticae</i> on pitkään tutkittu lähinnä maatalousaiheisena ammattikirjallisuutena, mutta uudemmassa tutkimuksessa on alettu kiinnostua teoksen dialogimuodosta, syvällisemmistä merkityksistä ja yhtäläisyyksiä filosofisen dialogin traditioon sekä satiiriin. Tekstissä esiintyvien eri henkilöhahmojen puheen sisällön ja etenkin kielellisten piirteiden eroihin ei kuitenkaan ole aiemmin kiinnitetty erityisempää huomiota.</p> <p>Puupankkiaineisto tekee mahdolliseksi määrällisen analyysin useista erilaisista kielellisistä piirteistä. Tässä työssä tarkastellaan erityisesti funktiosanojen frekvensseissä, <i>ut</i>-rakenteiden ominaisuuksissa ja alistuskonjunktoiden anastrofissa havaittuja eroja dialogin puhujien välillä. Funktiosanojen esiintymisfrekvenssejä tutkimalla voidaan saada vihjeitä esimerkiksi eri puhujien mahdollisesti käyttämistä sivulausetyypeistä. <i>Ut</i>-rakenteiden tarkempi tutkiminen puupankkiaineiston avulla on esimerkkinä yksityiskohtaisemmasta lauserakenteiden tarkastelusta. Anastrofi on aiemmassa tutkimuksessa esiin nostettu Varron kielelle tyypillinen.</p> <p>Puhujien välillä havaitut erot kielellisissä piirteissä näyttävät niin vahvoilta, että niiden tutkiminen perusteellisemmin vaikuttaa mielekkäältä. Erilaisten leksikaalisten ja syntaktisten piirteiden takana voi kuitenkin olla moninaisia syitä ja kielellisiä ilmiöitä, joten useimmissa tapauksissa menetelmien tässä tutkielmassa tuottamat havainnot toimivat lähinnä hyvinä vihjeinä perusteellisemmän tutkimuksen kohteiksi. Joka tapauksessa käytetyt digitaaliset menetelmät ja erityisesti puupankkien analysointi osoittautuvat tutkimuskysymyksen kannalta hyödyllisiksi ja käyttökelpoisiksi. Lisäksi tämänkaltaisen aineiston yleiskuvan tarkastelu parhaimmillaan tuottaa uusia havaintoja sellaisista erityispiirteistä, joihin ei välttämättä muuten tulisi kiinnittäneeksi huomiota.</p>		
Avainsanat – Nyckelord – Keywords dialogi, korpuslingvistiikka, latinan kieli, lemmafrekvenssi, puupankki, syntaksi		
Säilytyspaikka – Förvaringställe – Where deposited ethesis.helsinki.fi		
Muita tietoja – Övriga uppgifter – Additional information		

SISÄLTÖ

1	Johdanto.....	3
2	Marcus Terentius Varro ja Res Rusticae.....	5
2.1	<i>Varro grammaatikkona, antikvaarina ja kirjailijana.....</i>	<i>5</i>
	Antiikin kirjallisuuden kokoaja.....	6
	Varron oma tuotanto	6
	Varron tuotannon säilyminen ja vaikutus	8
2.2	<i>Res Rusticae</i>	<i>8</i>
	Omistus ja ajoitus.....	9
	Varron esikuvat ja lähteet	10
	Teoksen jälkivaikutus	11
3	Varron kielen erityispiirteistä ja aiemmasta tutkimuksesta	13
3.1	<i>Res Rusticaen dialogimuoto ja tyylilaji.....</i>	<i>13</i>
3.2	<i>Varron kielen erityispiirteitä.....</i>	<i>15</i>
	Ilmaisutyyli	16
	Viimeistelemättömyys	17
	Varrolle tunnusomaiset rakenteet.....	18
4	Aineisto ja tekstin prosessointi.....	21
4.1	<i>Käsikirjoitukset ja tekstin rekonstruktio.....</i>	<i>21</i>
4.2	<i>Digitaalinen aineisto</i>	<i>23</i>
	Repliikkien erottelu ja puhujien tunnistus	24
	Esikäsittely, virkejako, saneistus ja annotointi	24
4.3	<i>Tietotekniset apuvälineet ja tietokanta.....</i>	<i>26</i>
5	Metodologia.....	30
5.1	<i>Frekvenssianalyysi</i>	<i>30</i>
5.2	<i>Dependenssielioppi ja korpusten kielellinen annotointi</i>	<i>31</i>
	Suorat pääsana-määritesuhteet vs. siltarakenteet	34
	Rinnasteiset ilmaukset.....	35
	Ellipsi	36
	Kopula ja apuverbi	37
	Johtolauseet (parataksi).....	38
	Appositio ja nominien määreet	39
	Enhanced Dependencies	40
	Ut-rakenteiden analysointi ja annotointi	40

6	Tulokset.....	43
6.1	<i>Tekstin rakenne dialogina</i>	43
6.2	<i>Funktiosanojen frekvenssejä</i>	46
	Alistuskonjunktiot.....	47
	Rinnastuskonjunktiot	48
	Prepositiot	49
	Pronominit.....	50
	Apuverbi ja kieltosana	51
6.3	<i>Puupankki.....</i>	52
	Automaattinen jäsennys	53
	Anastrofi.....	54
	Ut-rakenteiden ominaisuuksia.....	55
7	Yhteenveto ja johtopäätökset.....	58
	Kirjallisuusluettelo	59
	Liitteet	62
	<i>Liite 1: Näytteitä aineiston käsittelyssä käytettävistä tiedostoista ja työkaluista</i>	63
	<i>Liite 2: Annotoinnissa käytetyt leimat ja relaatiot (UD-standardi).....</i>	66
	<i>Liite 3: Annotoidut esimerkkivirheet kokonaisina</i>	69
	<i>Liite 4: Dialogien kulku: puheenvuorojen pituudet ja vaihtelu</i>	71
	<i>Liite 5: Yleisimpien lemموjen esiintymät.....</i>	74

1 JOHDANTO

Tarkastelen tässä tutkielmassa Marcus Terentius Varron (116-27 eaa) *Res Rusticae*¹ -teoksen dialogimuotoisen tekstin henkilöhahmojen kieltä. Keskeisenä tutkimuskysymyksenä on selvittää, onko teoksessa esiintyvien puhujien kielessä havaittavissa eroja esimerkiksi sanastossa, morfologiassa tai syntaksissa. Samalla pyrin arvioimaan eräiden tietokoneavusteisten menetelmien soveltuvuutta tällaisen tutkimuskysymyksen ja aineiston käsittelyyn. Tutkimusmenetelminä käytän sanaston lemmafrekvenssianalyysiä keskittyen erityisesti funktiosanoihin sekä syntaktista annotointia, joka pohjalta voidaan analysoida myös lauserakenteisiin liittyvien piirteiden esiintyvyyttä.

Res Rusticae nähtiin varhaisemmassa tutkimuksessa pitkään lähes yksinomaan ammattikirjallisuutena, joka asettuu maataloutta käsittelevien oppaiden pitkään jatkumoon. Kerronnan muotoa pidettiin enimmilläänkin lähinnä kurioositeettina ja yrityksenä tehdä kuivasta asiatekstistä lukijalle kevyempää ja kirjallisesti kiinnostavampaa. Sen sijaan viimeaikaisessa tutkimuskirjallisuudessa teoksen dialogimuoto ja mahdolliset maatalousohjeita syvällisemmät merkitykset ovat nousseet keskeisiksi kiinnostuksen kohteiksi. Erityisesti yhtäläisyyksiä filosofisen dialogin traditioon sekä satiiriin on tutkittu monesta näkökulmasta.² Tietääkseni tekstissä esiintyvien eri henkilöhahmojen puheen sisältöön ja etenkin kielellisiin piirteisiin ei kuitenkaan ole vielä kiinnitetty erityisempää huomiota.

Dialogin puhujien kielessä voi olla eroja monesta eri syystä. Ilmeisimpiä tämän tekstin kohdalla ovat sanastolliset erot, jotka johtuvat yksinkertaisesti siitä, että päähenkilöt esiintyvät tietyn aihealueen asiantuntijoina, josta pääasiallisesti puhuvat. Olen tässä työssä rajannut tällaiset ammattisanastoon liittyvät seikat tarkastelun ulkopuolelle. Varron tiedetään käyttäneen kirjoitustyössään hyväkseen laajoja muistiinpanoja, ja tekstin tekniset osat ja ohjeet perustuvat suurelta osin antiikin varhaisempaan maatalouskirjallisuuteen. Koska puhujilla on omat erikoisalansa, on siis mahdollista sekin, että lähdeaineistossa käytetty kieli heijastuu myös *Res Rusticae* tekstissä. Varro on saattanut myös tarkoituksellisesti pyrkiä ilmentämään puhujien sosiaalista asemaa ja taustaa näiden käyttämässä kielessä. Tällaiset erot

¹ Käytän Varron kolmesta kirjasta tässä nimitystä *Res Rusticae*. Varsinkin uudemmassa englanninkielisessä tutkimuskirjallisuudessa teoksesta käytetään useimmiten nimitystä *De re rustica*. Käsikirjoituksissa ja varhaisissa editioissa (jotka sisältävät myös Caton ja Columellan maatalousaiheiset kirjat) näihin viitataan yleensä otsikolla *Rerum rusticarum libri tres* tai *De re rustica*. Jälimmäinen on myös Columellan kirjoista säännönmukaisesti käytetty nimitys.

² Nelsestuen 2015 on kattava esitys viimeaikaisen tutkimuksen kohteista ja tuloksista.

ovat luultavimmin lauserakenteisiin ja muoto-oppiin liittyviä, ja juuri tämän takia tarkastelen erityisesti tekstissä esiintyviä funktiosanoja ja syntaktisia piirteitä.

Tutkimalla funktiosanojen esiintymisfrekvenssejä voidaan saada osviittaa esimerkiksi eri puhujien käyttämistä sivulausetyypeistä. Vaikka tällainen tarkastelu on melko karkeaa ja edellyttää tarkempaa analyysiä pidemmälle menevien johtopäätösten tekemiseksi, sen etuna on, että koko kolme kirjaa kattava aineisto voidaan analysoida tietokoneavusteisia menetelmiä käyttäen suhteellisen vaivattomasti. Frekvenssianalyysin tulokset osoittavat, että dialogin puhujien käyttämässä kielessä tosiaan on systemaattisia eroavaisuuksia.

Lauserakenteita analysoida syntaktisen annotoinnin pohjalta tuotetusta puupankkiaineistosta. Tavoitteenani on tässä yhteydessä esittää annotoinnissa noudatettavat periaatteet ja osoittaa esimerkein lähestymistavan hyödyllisyys tutkimuskysymyksen kannalta. Lisäksi tarkastelen tietokoneavusteisesti toteutetun, koko tekstikorpuksen kattavan syntaksin automatisoidun annotoinnin käyttökelpoisuutta tekstin lauserakenteiden tutkimiseen.

Työn rakenne:

Seuraavassa luvussa esittelen taustaksi käsiteltävän tekstin ja sen kirjoittajan asemaa antiikin kirjallisuudessa ja erityisesti maatalousaiheisen ammattikirjallisuuden jatkumossa. Kolmas luku käsittelee Res Rusticaen kerronnan muotoa ja Varron kielen erityispiirteitä tähänastisen tutkimuskirjallisuuden valossa. Neljännessä luvussa kuvaan tutkittavan korpuksen tekstihistorian sekä tässä työssä käytetyn digitaalisen aineiston ja sen teknisen käsittelyn varsinaista analyysiä varten. Työssä käytettyjä kielitieteellisiä metodeja, erityisesti lemmafrekvenssianalyysiä ja syntaktisen annotoinnin periaatteita esitellään viidennessä luvussa. Kuudennessa luvussa esitän analyysin tulokset, ja seitsemäs luku sisältää johtopäätöksiä ja yhteenvedon.

2 MARCUS TERENTIUS VARRO JA RES RUSTICAE

Tämä luku taustoittaa tässä työssä käsiteltävää tekstiä ja sen kirjoittajaa antiikin kirjallisuuden kontekstissa. Aluksi esittelen lyhyesti Marcus Terentius Varron (116–27 eaa.) elämää ja erityisesti monitahoista ja keskeistä asemaa Rooman ja yleisemminkin antiikin kirjallisuuden välittäjänä varhaisvaiheista nykyaikaan. Tämä osa käsittää Varron eri roolit häntä edeltävän kirjallisuuden kokoajana ja tallentajana, itsenäisenä kirjallisuuden tuottajana sekä jälkivaikutusten lähteenä. Tämän jälkeen havainnollistan näitä rooleja käytännössä tässä työssä tutkittavassa tekstissä, Varron ainoassa lähes kokonaan säilyneessä teoksessa, *Res Rustica*essa.

2.1 Varro grammatikkona, antikvaarina ja kirjailijana

Vuonna 116 eaa. syntynyt Varro oli kotoisin sabiinilaisalueella sijaitsevasta Reaten (nyk. Rieti) kaupungista. Varron grammatikon ja antikvaarin kutsumukset heräsivät jo nuorena, kun hänellä oli opettajanaan Lucius Aelius Stilo, jota Cicero kuvaa erittäin oppineeksi latinan ja kreikan kielen sekä Rooman historian tuntijaksi.³ Lisäksi Varro kuului Titus Pomponius Atticuksen, traagisen runoilijan ja kirjankustantajan ja kulttuurinedistäjän, lähipiiriin. Varro jatkoi opintojaan filosofian alalla viettäen vuodet 84–82 Ateenassa opettajanaan Antiokhos Askalonilainen, joka näyttää kuuluneen aikansa tärkeimpiin ajattelijoihin.⁴

Varron omasta tuotannosta ja aikalaisten, esimerkiksi Plinius vanhemman, todistuksesta saadaan viitteitä tämän monitietäjän työskentelytavoista ja -metodeista. Ehkä leimallisinta on hämmästyttävä työskentelyn nopeus, joka selittää suurelta osin sekä kirjallisen tuotannon että muun toiminnan valtavan määrän. Toisaalta nopea työtahti selittää myös huolittelemattoman lopputuloksen – Varro ei ilmeisesti juurikaan hukannut aikaa omien kirjoitustensa oikolukuun. Asioiden käsittely noudattaa Varron tunnetussa tuotannossa yleensä samaa kaavaa: ihmisistä (*de hominibus*), paikoista (*de locis*), ajanjaksoista (*de temporibus*) ja asioista (*de rebus*). Pytagoralaisen numerologian vaikutuksesta Varro on suorastaan jumiutunut jakamaan käsittelemänsä asiat tiettyihin määriin lokeroita ja näiden alaluokkia. Kerran kerätty materiaali voitiin käyttää useita kertoja uudelleen: Esimerkiksi tutkielmatyyppisen teoksen johtopäätöskappaleen muodon Varro esitti ilmeisesti ainakin neljässä eri teoksessa. Varro teki runsaasti muistiinpanoja, jotka hän myös luokitteli ja

³ Cic. *Brut.* 205.

⁴ Heurgon 1978, x-xiii.

järjesteli. Tosin näidenkin käytössä näkyy, ettei muistiinpanoja ole aina käytetty tekstin viimeistelyyn tai yksityiskohtien tarkastamiseen.⁵

Antiikin kirjallisuuden kokoaja

Varro kokosi antiikin kirjallisuutta myös melko kirjaimellisessa mielessä. Caesar antoi vuonna 46 eKr Varron tehtäväksi Rooman ensimmäisen julkisen kirjaston perustamisen. Tehtävänä olisi ollut koota, järjestää ja saattaa yleisön käyttöön valikoimiltaan mahdollisimman kattavat latinalaiset ja kreikkalaiset kirjastot, mutta vuoden 44 maaliskuun *Idus*-päivän tapahtumat pysäyttivät hankkeen. Vuonna 43 Varro joutui Antoniuksen proskriptiolistalle, mutta säilytti henkensä. Varron julkinen ura loppui tähän, mutta kirjallisia töitään hän jatkoi vielä 15 vuoden ajan, kuolemaansa saakka.⁶

Kirjastohankeen kaatumisesta huolimatta Varro on vaikuttanut Rooman kirjallisuuden säilymiseen ja valikoitumiseen jälkipolville. Aleksandrialaista filologien esimerkkiä seuraten roomalaiset oppineet ottivat tehtäväkseen vanhojen kirjallisten tekstien alkuperäisyyden määrittämisen. Tässä työssä Varro oli yksi ensimmäisiä roomalaisia ja hän esimerkiksi määritteli Plautuksen tuotannon auktoritatiivisen kaanonin, jonka 21 näytelmää ovat ainoat keskiaikaiseen käsikirjoitustraditioon säilyneet.⁷

Varron oma tuotanto

Varron toiminnan suurin merkitys antiikin kirjalliselle traditiolle välittyi epäilemättä hänen oman tuotantonsa kautta. Varron kirjallinen tuotteliaisuus on sekä määränsä että monipuolisuutensa suhteen vertaansa vailla. Yhdeksänkymmenen elinvuotensa aikana hän tuotti 74 teosta – yli 600 kirjan verran tekstiä – monilta eri aloilta. Tästä valtaisasta määrästä on kokonaisuudessaan säilynyt vain kuusi *De lingua Latina* -teoksen yhteensä 25 kirjasta sekä tämän tutkielman aiheena oleva, kolme kirjaa sisältävä maanviljelysopas *Res rusticae*.⁸

Varron tuotanto oli paitsi runsas, myös aikalaisten suuresti arvostama. Cicero ylistää kirjeessään Varroa, joka oli ”palauttanut roomalaisille näiden oman historian”.⁹ Quintilianus sanoo Varron olevan *vir Romanorum eruditissimus*¹⁰ ja Plutarkhos samaan tapaan *ἄνδρα Ῥωμαίων ἐν ἱστορίᾳ βιβλιακώτατον*¹¹. Maine oli yhtä vankka humanistienkin keskuudessa, ja

⁵ Horsfall 1982, 286-7.

⁶ esim. Riposati 1976, 61–66; Heurgon 1978, XIII–XV.

⁷ Kenney 1982, 27-8.

⁸ Heurgon 1978, XVI–XIX; Horsfall 1982, 286-7.

⁹ Cic. Ac. 1,9.

¹⁰ Quint. Inst. 10,1,95.

¹¹ Plut. Rom. 12,3.

Petrarca kutsuukin Varroa nimellä *il terzo gran lume di Roma*, ”Rooman kolmas suuri valo” nostaen tämän näin Ciceron ja Vergiliuksen rinnalle.¹²

Varron mielenkiinnon kohteet vaihtelivat ajan kuluessa ja kirjoitustyö kohdistui yleensä melko rajattuun aihepiiriin kerrallaan. Asioiden luokitteluun mieltynyt Varro voitaisiin itse lokeroida tuotantonsa perusteella ”alatyypeiksi”.¹³ Varhaisimpaan tuotantoon kuluu *Runoilija-Varro*, jolta on säilynyt noin 600 fragmenttia *Menippolaissatiireita*. Nämä vaihteleviin mittoihin seipitetyt, proosajaksojen katkomat Varron nuoruuden teokset arvostettiin aikanaan taiteellisen omaperäisyyden mestariteokseksi. Pääasiassa myöhäisimpään tuotantoon kuulunut *Historioitsija*-, *oikeustieteilijä*- ja *teologi-Varro* selvitti mitä moninaisimpien uskonnollisten ja valtiollisten tapojen ja menojen taustoja. Juuri Varro oli määrittänyt Rooman traditionaalisen perustamisvuoden ja hänellä oli mahdollisesti rooli myös Caesarin kalenteriuudistuksen takana. Ilmeisesti ainoa läpi elämän säilynyt kiinnostuksen kohde oli latinankielen luonne ja historia. Tämän *kielitieteilijä-Varron* varhaisin teos *De antiquitate litterarum* on 80-luvun alusta, *Disciplinae I* ja *De grammatica* yli puoli vuosisataa myöhäisempiä. Tästä tuotannosta tunnemme osan *De Lingua Latinan* kuuden säilyneen kirjan muodossa.¹⁴

Varro ei välittänyt aikaisempaa kirjallisuutta pelkästään asiasisällön osalta. Edellä mainitut *Menippolaissatiirit*, joita Varro kirjoitti 150, jäljittelevät kolmannella vuosisadalla eKr. eläneen syyrialaisen kyynikkofilosofi ja satiirikirjailija Menippoksen proosaa ja runomittaa sekä huumoria ja moralisointia yhdisteleviä kirjoituksia. Varron versioissa yhdistyvät menestyksekkäästi Vähän-Aasian elämänilo ja sabiinilainen ajatusten eheys. Jälkimmäiseen liittyy Varron mieltymys filosofiaan, joka ilmenee myös läpi koko kirjallisen tuotannon.

Myöhäisemmässä tuotannossa erottuvat erillisinä varsinkin historioitsijan ja kielitieteilijän roolit. Historian osalta käsittelyssä olivat erityisesti Rooman historia ja uskonnon historia. Näistä Varro kirjoitti 45 kirjaa käsittävän *De antiquitates* teoksen, joka jakautui kahteen osaan: *Res humanae* (vuoden 56 tienoilla) ja *Res divinae* (v. 47). Tämän suurteoksen lisäksi tuotantoon kuului useita pienempiä historiateoksia, kuten *De familiis Troianis* ja *De vita populi Romani* sekä oikeushistoriaa käsittelevä, 15 kirjan *De iure civili*. Varron toimeliaisuuden taukoamattomuutta ja perusteellisuutta kaikkein suurimmassa määrin ilmentää hänen kielitieteellinen tuotantonsa. Tunnetuin ja merkittävin

¹² Della Corte 1954, 5.

¹³ esim. Heurgon 1978, VIII–X.

¹⁴ Riposati 1976, 83; Horsfall 1982, 286–7.

osa tästä on 25 kirjaa käsittänyt, Cicerolle omistettu *De lingua Latina*, mutta tämän lisäksi Varro kirjoitti mm. teatterin historiasta.¹⁵

Varron tuotannon säilyminen ja vaikutus

Jo keisariajalla huomattava osa Varron laajaa tuotantoa tunnettiin enää lyhennelmien välityksellä.¹⁶ Esimerkiksi *Menippolaissatiirien* säilyneistä katkelmista 95 prosenttia on sitaatteja 400- ja 500-lukujen taitteessa vaikuttaneen Nonius Marcelluksen *De compendiosa doctrina* -teoksessa.¹⁷ Alkuperäisen tuotannon lähes täydellinen katoaminen voidaan kuitenkin nähdä osituksena Varron suuresta vaikutuksesta antiikin myöhempään kirjallisuuteen: hänen töitään käytettiin myöhempien kirjoittajien lähteenä niin runsaasti, että alkuperäiset tekstit kävivät tarpeettomiksi säilyttää. Varron vaikutus näkyi jo Vergiliuksen ja Ovidiuksen tuotannoissa, myöhemmin Columellalla ja Suetoniuksella ja ulottui vielä myöhäisantiikkiin esimerkiksi Ausoniuksen ja Isidoruksen teksteihin.¹⁸

Tietokirjallisuuden vaikutukset ulottuvat myös muualle yhteiskuntaan. Keskiaikaisen yliopisto-opetuksen vapaat taidot, *trivium* (kielioppi, retoriikka, dialektiikka) ja *quadrivium* (aritmetiikka, geometria, tähtitiede, musiikin teoria), olivat peräisin Varron myöhäisvuosien *Disciplinae*-teoksesta. Jälkiä tästä Varron systematisoinnista näkyy vielä nykyisinkin yliopistojen opinto-ohjelmissa.¹⁹

2.2 *Res Rusticae*

Vaikka maatalousopas saattaa näyttää Varron tuotannon joukossa lähinnä pieneltä kuriositeetilta, juuri sen säilyminen kokonaisuudessaan ei ole pelkkää sattumaa. Tällaisella teoksella oli välitöntä käyttöarvoa myöhäisantiikkiin asti, ja maataloutta käsitteleviä kirjoja yleensäkin kopioitiin ja käytettiin uusien oppaiden lähteenä varsin ahkerasti. Teos itse ja sen käyttöhistoria ovat myös havainnollisia esimerkkejä Varron roolista antikin kirjallisuuden jatkumossa.

Teos koostuu kolmesta kirjasta, joista ensimmäinen käsittelee maanviljelystä (*De agricultura*), toinen karjankasvatusta (*De re pecuaria*) ja kolmas maatalon muita eläimiä ja tuotteita sekä näiden kaupallista hyödyntämistä (*De villatica pastione*). Teos on kirjoitettu dialogimuotoon, joka värittää ja elävöittää muutoin teknisen aiheen käsittelyä. Kertomuksen

¹⁵ Heurgon 1978, XVI-XIX; Horsfall 1982, 288-90.

¹⁶ Castrén & Pietlä-Castrén 2000, 616-7.

¹⁷ Cèbe, 1972, VIII.

¹⁸ Horsfall 1982, 286.

¹⁹ *ibid.*

atmosfääri on leimallisesti roomalainen, mikä ei ole sattumaa. Roomalaisille maanviljelys oli tunnepitoinen, läheisesti kotiseuturakkauteen liittyvä aihe, jota hyödynnettiin runsaasti myös runoudessa.

Ensimmäistä kirjaa kirjoittaessaan Varro ei näytä ajatelleen suurmaanomistajia vaan vaatimattomammin resurssein toimivia maanviljelijöitä, joiden tilat esitetään useassa kohdin vastakkain rikkaiden ja hyvin varustettujen maaseututalojen kanssa. Toinen kirja on selvästikin tarkoitettu Varron itsensä kaltaisille suurten karjatilojen omistajille. Kolmas kirja menee vielä astetta pidemmälle esittäen rikkaiden maanomistajien keskusteluja moninaisista ylellisyystuotteista.

Varro jakaa aiheensa systemaattisesti suorastaan pedanttisella huolellisuudella pääluokkiin, alaluokkiin ja edelleen alaluokkien alaluokkiin ja selkiyttää esitystään tiivistelmänomaisin johdannoin ja yhteenvedoin. Maanviljelys aiheena on kuitenkin systematisoinnin kannalta vähemmän kiitollinen, minkä seurauksena Varron luokitteluyritys vaikuttaa usein keinotekoiselta, vain luokittelun takia tehdyltä.²⁰

Omistus ja ajoitus

Ensimmäisen kirjan alussa Varro omistaa teoksensa nuorelle vaimolleen Fundanialle ja selittää kirjoittavansa kiireessä, koska on jo 80-vuotias. Fundania on ostanut maatilan, jonka toivoo tulevan tuottavaksi ja on pyytänyt mieheltään neuvoja tämän päämäärän saavuttamiseksi. Teoksen avulla Varro toivoo voivansa olla avuksi myös kuolemansa jälkeen:

*Quare, quoniam emisti fundum, quem bene colendo fructuosum cum facere velis, meque ut id mihi habeam curare roges, experiar; et non solum, ut ipse quoad vivam, quid fieri oporteat ut te moneam, sed etiam post mortem.*²¹

Niinpä, kun kerran olet ostanut maatilan, ja koska haluat hyvin viljelemällä tehdä siitä tuottoisan, ja kun pyydät minua ottamaan asian hoitaakseni, yritän: eikä vain, että opetan kuinka pitää toimia, kun itse olen elossa vaan myös kuolemani jälkeen.

Omistuskirjoitus on hämmentänyt tutkijoita kahdesta syystä. Ensinnäkin vain ensimmäinen kirja vastaa Fundanialle annettua lupautta ja soveltuu ohjeeksi tämän hankkiman tilan hoidossa; toinen kirja käsittelee karjanhoitoa ja kolmas erilaisia ylellisyystuotteita. Toiseksi kaksi jälkimmäistä kirjaa sisältävät molemmat omat omistuksensa kokonaan eri henkilöille. Toisessa kirjassa on omistus Turranius Nigerille ja kolmannessa Quintus Pinnaukselle.

²⁰ Skydsgaard 1968, 9–10.

²¹ Varro *Rust.* 1,1,1.

Ongelmaan on tarjottu ratkaisuksi niin Varron hajamielisyyttä tai suunnitelman muutosta kuin käsikirjoitusten kopiointivirheitäkin.²²

Ensimmäisen kirjan johdannon ja kirjojen sisällöllisen vertailun pohjalta on esitetty, että kolme kirjaa olisikin kirjoitettu eri ajankohtina ja vain ensimmäinen olisi kuulunut alkuperäiseen suunnitelmaan Fundaniaa varten kirjoitetusta opastuksesta.²³ Nykyisin pidetään kuitenkin melko varmana, että kaikki kolme kirjaa on kirjoitettu peräjälkeen ja julkaistu vuonna 37 eaa.

Varron esikuvat ja lähteet

Caton *De agri cultura* ja Varron *Res Rusticae* kirjoittamisen väliin mahtuu yli sadan vuoden myrskyisä ajanjakso, jonka päätepiirteenä oli kansalaissota ja Rooman tasavallan ajan loppuminen. Maanviljelyn saralla tämä ajanjakso merkitsi itsenäisten pienviljelijöiden väistymistä tehokkaamman viljelyteknologian ja karjankasvatuksen tieltä. Varro oli itsekin menestyksekkäs karjankasvattaja ja maanomistaja.²⁴ Varron lähestymistapa poikkeaa merkittävästi Catosta. Catolle, ja Varron omana aikana Cicerolle, maanviljelys oli etupäässä taito, jonka kuka tahansa saattoi ilman erityiskoulutusta hallita, ja maalaiselämä nähtiin osana *otiumia*. Varrolle, epäilemättä Scrofaa (ks. alla) seuraten, maanviljelys kuului *negotiumin*, liiketoiminnan, piiriin. Varron jälkeen Columella meni vielä pidemmälle pitäen sitä ammattina (*occupatio*).²⁵

Varro kertoo esipuheessaan mainitsevansa sekä kreikkalaisia että roomalaisia auktoreita, joiden puoleen Fundania voisi kääntyä tarvitessaan lisätietoja. Listalla on 39 kreikkalaista kirjoittajaa sekä karthagolainen Mago, muttei yhtään roomalaista. Columellan mukaan Magon 28 maanviljelystä käsittelevää kirjaa pidettiin niin merkittävinä, että ne käännettiin latinaksi senaatin määräyksestä.²⁶

Caton jälkeistä maanviljelyskirjallisuutta ennen Varron teoksia ei ole säilynyt, mutta Varron itsensä ja muiden antiikin kirjoittajien todistuksen perusteella merkittäviä nimiä alalla olivat Saserna nuorempi ja vanhempi sekä Varron dialogeissakin keskeistä osaa näyttelevä Tremelius Scrofa.²⁷ Juuri Scrofa näyttää ensimmäisenä käsitelleen aihetta tieteellisemmän ja teknisemmän lähestymistavan kautta teollisuudenhaarana hyläten

²² esim. White 1973, 493.

²³ esim. Martin 1971, 224–233.

²⁴ White 1973, 458.

²⁵ Martin 1995, 90.

²⁶ Colum. 1,1,13.

²⁷ White 1973, 459–461.

perinteisen näkökulman, jossa maanviljelys käsitettiin lähinnä talonpoikaisena elämäntapana.²⁸ Vaikka Scrofa oli ilmeisesti Varron välitön edeltäjä maatalouskirjailijana, häntä ei kuitenkaan ole mainittu johdannon listassa.

Aikaisemmin Varroa pidettiin huolimattomana kirjoittajana, joka oli siteerannut lähteitään huolimattomasti ja jopa ymmärtänyt ne väärin.²⁹ Varro itse sanoo toistavansa maanviljelyksen alalta hiljattain käydyt keskustelut.³⁰ Uudempi tutkimus on kuitenkin osoittanut, ettei Varro ollut huolimaton ja ymmärtämätön jäljittelijä vaan itsenäinen kirjoittaja, joka käytti taitavasti hyväkseen aikaisempaa materiaalia muokaten sitä lukijan kannalta ymmärrettävämpään muotoon sovittamalla sen Italian oloihin ja päivittämällä vanhentuneita tietoja. Tosin kiireen jälki näkyy tekstissä paikoitellen. On myös helppo havaita, milloin Varro on kirjoittanut itselleen tutuista aiheista omaa kokemustaan hyödyntäen ja milloin pelkästään auktoriteetteihin nojaten. Jälkimmäisissä tapauksissa tosiaan esiintyy epätäsmällisyyksiä ja suoranaisia virhetulkintoja.³¹

Teoksen jälkivaikutus

Res rusticae antoi mahdollisesti vaikutteita jo Vergiliuksen (70–19 eKr.) *Georgicaan*, jonka kirjoittaminen ajoittuu samoihin aikoihin kuin Varron maatalousoppaan julkaisu, sekä Hyginuksen (60 eKr.–10 jKr.) ja Celsuksen (n. 20 eKr.–50 jKr.) maatalousaiheisiin kirjoituksiin. Varmuudella voidaan sanoa, että perusteellisen, 12 kirjaa käsittävän *De re rustica* –teoksen kirjoittanut maatalouskirjailija Columella (4–n. 70 jKr.) sekä oppinut ja historioitsija Plinius vanhempi (23–79 jKr.) tunsivat Varron maatalousaiheiset kirjoitukset hyvin. Vaikka Columella ei mainitsekaan bibliografiassaan Varroa muuten kuin maanviljelyksen kaunopuheistajana (*”eloquentem reddidit”*), on aivan ilmeistä, että hänen teoksessaan Varrolta ovat peräisin kreikkalainen kirjallisuusluettelo ja monet maalaiselämän konkreettiset yksityiskohdat.³² Plinius siteeraa Varroa luonnonhistoriassaan (*Naturalis historia*) nimeltä mainiten lukuisissa kohdissa. Koko keisariajan ja myöhäisantiikkiin asti Varron *Res rustica* on useilla kirjailijoilla sitaatteja, jotka ovat merkittävä lisä

²⁸ Martin 1971, 247–248.

²⁹ esim. Gentili 1903.

³⁰ Varro *Rust.* 1,7.

³¹ Skydsgaard 1968, 64–88.

³² Ensimmäisessä kriittisessä editiossa *Res Rusticae* (Teubner-editio vuodelta 1891) Heinrich Keil osoittaa 17 tällaista kohtaa.

käsitteistötraditioon.³³ Erityisesti edellä jo mainittu Noniuksen *De compendiosa doctrina* sisältää 42 suoraan *Res Rusticae* otettua esimerkkiä.³⁴

Varron nimi esiintyy taajaan kreikankielisessä, 900-luvun puolivälissä ilmestyneessä kokoelmassa maatalousaiheisia kirjoituksia, jotka on omistettu keisari Konstantinos VII Porfyrogennetokselle ja tunnetaan nimellä *Geoponica*. Tässä teoksessa Varro näyttää kuitenkin edustavan varhaisemman ajan maatalouskirjailijoita yleensä ja saa nimiinsä myös ilmeisesti Columellalta peräisin olevia katkelmia, mutta erityisesti *Res Rusticae* toisesta ja kolmannesta kirjasta on *Geoponicassa* sisällöltään ilmeisen muuttumattomina säilyneitä otteita. Osa *Geoponica* käännettiin myös arabiaksi, minkä lisäksi on mahdollista, että Andalusianssa 600- ja 700-lukujen taitteessa vaikuttaneen Ibn-al-Awwamin maatalouskirja sisältää myös Varrolta suoraan saatuja lainauksia.³⁵

³³ esim. Aulus Gellius (n. 125–180), Charisius ja Diomedes (vaikuttivat 300-luvun lopulla), Macrobius (300- ja 400-lukujen taite) ja Priscianus (400- ja 500-lukujen taite).

³⁴ Heurgon 1978, L-LIII.

³⁵ Rodgers 1978; Heurgon 1978, LIII-LV.

3 VARRON KIELEN ERITYISPIIRTEISTÄ JA AIEMMASTA TUTKIMUKSESTA

3.1 *Res Rusticaen dialogimuoto ja tyyli*

Kuten edellä on esitetty, Varron *Res Rusticae* kytkeytyy käsikirjoitustradition ja aiheensa puolesta antiikin ammattikirjallisuuteen, erityisesti sitä edeltäneeseen Caton *De Agri Cultura*an ja seuranneeseen Columellan *De re rustica*an. Muodoltaan, tyyliltään ja koko sisällöltään *Res Rusticae* ei kuitenkaan rajaudu maanviljelyksen oppikirjaksi. Dialogimuotoinen kertomus sekä dialogien huolellisesti valitut näyttämöt ja värikkäät henkilöt vertautuvat Platonin, Ksenofonin ja etenkin Ciceron filosofisiin dialogeihin. Runsas sanaleikkien viljely, ironia ja dialogin henkilöhahmojen yllätyskohtaukset puolestaan liittävät tekstin satiiriseen perinteeseen.³⁶ Varro myös höystää kerrontaansa kuivakkaalla huumorillaan ja nokkelalla sanankäytöllään. Myös edellä mainittu läpi Varron elämän kestänyt intohimo grammatiikkaan antaa oman sävynsä tekstile, jossa on siellä täällä etymologisia selityksiä ja kielellistä analyysiä.³⁷

Varron kerronnan muoto oli *Res Rusticaen* tekstin tutkimuksessa pitkään sivuosassa, jos edes siinä. Parhaimmillaankin dialogimuoto nähtiin kirjoittajan keinona – ja kömpelölkönä sellaisena – tehdä kuivasta asiatekstistä lukijalle kevyempää ja kirjallisesti kiinnostavampaa.³⁸ Lukuun ottamatta yhtä 1950-luvulla ilmestynyttä väitöskirjaa³⁹, tekstin muotoon ja tyyliin on alettu kiinnittää huomiota vasta viimeisen parinkymmen vuoden aikana.⁴⁰

Grant A. Nelsestuen, joka on hiljattain analysoinut *Res Rusticaen* dialogeja ja kattavasti esitellyt niiden viimeaikaista tutkimusta, esittää teoksen edustavan aivan omaa lajityyppiään, jossa yhdistyy piirteitä filosofisista teksteistä, satiirista ja ammattikirjallisuudesta: Ilmeisin yhteys filosofiseen kirjallisuuteen on tekstin dialogimuoto itsessään, kuten edellä on jo mainittu.⁴¹ Varro on ilmeisesti pitänyt mallinaan Ciceron *De Oratore*a, sillä dialogin muodossa ja henkilöissä on havaittavissa selviä yhtäläisyyksiä näiden teosten välillä ja myös kysymyksenasettelu⁴², ”onko maanviljelys *ars*”, on samankaltainen.⁴³ Varron dialogi ei kuitenkaan typisty esikuviansa, ennen muuta Ciceron, pintapuoliseen ja ehkä parodioivaan

³⁶ Nelsestuen 2015, 9.

³⁷ Trevizam 2011; 2012.

³⁸ Esim White 1973, 488–94.

³⁹ Heisterhagen, 1952.

⁴⁰ Nelsestuen 2015 mainitsee ensimmäisinä tätä lähestymistapaa edustavina artikkeleina Purcell 1995 ja Green 1997, monografioina Diederich 2007 ja Kronenberg 2009.

⁴¹ Nelsestuen 2015, 9–30.

⁴² Varro *Rust.* 1,3,1; Cic. *Orat.* 22,102.

⁴³ Martin 1995, 81–84.

jäljittelyyn. Diaologeissa on sekä pää- että sivuhenkilöitä kuten Cicerolla, mutta Varrolla sivuhenkilöitä on useampi kuin Cicerolla tavallisesti ja näiden rooli on jonkin verran merkittävämpi. Ciceron dialogeissa osallistujat ovat yleensä tunnistettavissa todellisiksi aikalaisiksi tai historiallisiksi (merkki)henkilöiksi, kun taas Varrolla on mukana tuntemattomia tai ilmeisen keksittyjä hahmoja, jotka näyttävät olevan valittu mukaan lähinnä aiheeseen sopivan nimensä perusteella.⁴⁴

Ensimmäisessä kirjassa on kaksi selkeää päähenkilöä (Gnaeus Tremellius Scrofa ja Gaius Licinius Stolo) ja neljä tai viisi sivuhenkilöä (Varro itse, Gaius Fundanius, Publius Agrasius ja Gaius Agrius sekä kokoonkutsuja Lucius Fundilius, joka ei osallistu keskusteluun). Näistä Scrofa on antiikin aikana tunnettu maatalouskirjailija. Todellisia historiallisia henkilöitä ovat ilmeisesti myös Stolo ja Varron appi Fundanius. Toisessa kirjassa puheenvuorot ja puhe aika jakautuvat tasaisemmin ja kaikilla seitsemällä henkilöllä on dialogissa merkittävä rooli. Kirjassa esiintyvät jälleen Scrofa ja Varro itse sekä Titus Pomponius Atticus, Lucius Cossinius, Quintus Lucienus ja Vaccius. Atticus ja Lucienus ovat tunnettuja historiallisia henkilöitä, Vaccius (”Lehmänen” tms.) ilmeisen keksitty, ja Cossiniuksen henkilöllisyydestä ei ole varmuutta. Kolmannessa kirjassa pääpuhujina ovat Varro, Quintus Axius, Appius Claudius Pulcher ja Cornelius Merula, jotka kaikki ovat muista lähteistä tunnettuja ja ilmeisesti todellisuudessaakin Varron tuttavapiiriin kuuluneita. Tässä mielessä kolmas kirja on selkeimmin edellä mainitun, Ciceron suosiman ”Aristoteleen mallin” mukainen. Hyvin pienessä roolissa sivuhenkilöinä esiintyvät Fircellius Pavo (”riikinkukko”) ja Minucius Pica (”harakka”) sekä kokonaan ilman puheenvuoroja Pantuleius Parra (”pöllö”) ja Marcus Petronius Passer (”varpunen”).

*

Tulkinnan kannalta merkittävin ja mahdollisesti hyvin harkittu piirre on dialogien sijoittaminen julkisiin tiloihin ja julkisten tapahtumien yhteyteen. Siinä missä Ciceron dialogeissa tyypillisesti keskustellaan valtion asioista filosofisesti maaseudun rauhassa, *Res Rustica*essa puhutaan kaupunkiympäristössä maanviljelyksestä. Tämä voi olla tietoisesti rakennettu peilikuva Ciceron dialogeihin, mutta korostaa toisaalta tekstin poliittisia ulottuvuuksia ja allegorioita, joissa maanviljelys linkittyy julkiseen elämään ja toimintaan.⁴⁵

⁴⁴ Cicero jakaa dialogit Aristoteen ja Heraclideen mallin (*Arstotelius mos*, *Heraclideus mos*) mukaisiin, josta ensin mainituissa kirjoittaja itse keskustelee aikalaisensa kanssa ja jälkimmäisessä henkilöt ja tapahtumat sijoittuvat menneisyyteen. Varron dialogeja on yritetty sovittaa tähän jakoon (esim. Nelsestuen 2016, Martin 1995, Heurgon 1978), mutta nähdäkseni tämä on keinotekoista, ja em. kirjoittajatkin myöntävät erottelun vaikeaksi.

⁴⁵ Nelsestuen 2015, 11–19.

Res Rusticaen satiiriin viittaavat piirteet eivät ole sinänsä yllättäviä, koska tyyllilaji oli Varrolle hyvin läheinen; kuten edellä on esitetty, hänen varhaisempaan tuotantoonsa kuului 150 Menippolaissatiiria. Sitä paitsi satiiri on yleensäkin tavallinen elementti filosofisissa dialogeissa. Satiirina tulkiten teoksessa esitetyt hyvät maatalouden käytännöt voidaan nähdä metaforana hyvälle kaupungin ja valtion asioiden hallinnolle.⁴⁶ Varrolle satiiri antaa myös keinon kritisoida perinnäisten tapojen ja moraalin rappiota teknisten maatalousohjeiden ohessa. Laajemmin katsottuna teoksen satiiriset elementit eivät rajoitu yhtäläisyyksiin varsinaisten satiirien kanssa, vaan sisältävät myös sellaisia keinoja kuin parodia, ironia, moralisointi, sanaleikit ja huumori. Etenkin sanaleikkejä on tekstissä runsaasti ja erityisen silmiinpistävästi dialogin henkilöiden nimissä.⁴⁷

Vaikka *Res Rusticaen* viimeaikaisessa tutkimuksessa kertomuksen dialogimuoto on noussut keskeiseen osaan, tekstissä esiintyvien eri henkilöhahmojen puheen sisältöön ja etenkin kielellisiin piirteisiin ei tietääkseni ole vielä kiinnitetty erityisempää huomiota. Aihetta sivuten Heurgon kyllä toteaa henkilöiden luonteenlaadun ja aseman heijastuvan näiden puheessa, mutta tarkastelu jää repliikkien sävystä syntyneiden vaikutelmien varaan.⁴⁸ Matheus Trevizam taas on hiljattain tutkinut Varron sanastoa ja kieltä etenkin verrattuna muihin saman aikakauden maatalousaiheisiin teoksiin, mutta päähuomio on pikemminkin tyyllilajien eroissa kuin Varron ilmaisun kielellisissä piirteissä sinänsä.⁴⁹

Erityisesti näytelmäkirjallisuudessa, ja myöhemmin esimerkiksi Petroniusken *Satyricon*-teoksessa, on tyypillistä, että eri henkilöhahmojen käyttämä kieli heijastelee näiden taustaa. Koska Varron teosta on vaikea verrata muuhun dialogimuotoiseen kirjallisuuteen, ei liene mahdotonta, että myös puhujien kielessä olisi eroja. Ylipäätään tutkimuskirjallisuus, jossa analysoidaan systemaattisesti Varron tuotannon kielellisiä piirteitä, on yllättävän niukkaa ottaen huomioon, kuinka paljon Varron tyyliä on etenkin vanhemmassa keskustelussa parjattu. Seuraavassa pyrin kokomaan yhteen keskeiset Varron tyyliä ja kieltä käsittelevät huomiot.

3.2 Varron kielen erityispiirteitä

Varron kieltä koskevaa keskustelua on jo aikalaiskommenteista lähtien hallinnut sangen vähättelevä sävy sekä yhtäältä kielellisten puutteiden ja toisaalta Varron ilmeisen

⁴⁶ Green 1997, 427–48.

⁴⁷ Nelsestuen 2015, 19–27.

⁴⁸ Heurgon 1978, XLVI.

⁴⁹ Esim. Trevizam 2011; 2013; 2014.

oppineisuuden välisen vastakkainasettelun kommentointi. Kielelliseen ilmaisuun ovat epäilemättä vaikuttaneet materiaalina käytetyt, kiireessä tehdyt muistiinpanot sekä dialogin puhekielenomaisuus yhdistettynä Varron mieltymykseen retoriikan keinoihin. Kun näiden lisäksi tekstissä on runsaasti anakoluutteja eli virkkeen alun ja lopun välisiä lauseopillis-loogisia yhteensopimattomuuksia, on jopa kokonaisten lauseiden merkityksen tulkinta paikoin vaikeaa.⁵⁰

Näin ollen Varron teosten lukemista on pidetty hyödyllisenä lähinnä niiden asiasisällön takia – ja kielestä huolimatta. Kielelliseltä kannalta tekstejä on pidetty lukemisen arvoisina korkeintaan varoittaviksi esimerkeiksi. Jopa Eric Laughton, joka muutoin osoittaa monia Varron kieleen kohdistuneita arvioita liioitellun negatiivisiksi, toteaa aihetta käsittelevän artikkelinsa *Observations on the Style of Varro* alussa:

*Ancient and modern scholars are so unanimous in their condemnation of Varro as a writer, that a study of his 'style' may seem to be valueless. [...] Nevertheless some attention to Varro's manner of writing will not be unprofitable. To be confronted with the clumsiness and complexity of a Varronian period—if that is what it should be called—is to have a heightened admiration for the clarity and elegance of Caesar and of Cicero.*⁵¹

Myöhemmin Laughton kuitenkin huomauttaa, että Varron kielen perusteellisempi analyysi on suureksi hyödyksi itse tekstien rekonstruktiossa ja auttaa tunnistamaan niiden vaikutuksia myöhäsisemmässä Rooman kirjallisuudessa, jossa Varroa käytettiin hyvin laajasti lähteenä, kuten jo edellä on todettu. Tämä vuonna 1960 julkaistu artikkeli on tietävästi edelleen uusin varsinaisesti kielellisiä piirteitä Varron teksteissä kattavasti käsittelevä tutkimus, ja seuraava esitys perustuu kokonaisuudessaan siihen.

Ilmaisutyyli

On väitetty, että Varron ilmaisussa hämmäntävällä tavalla tarkoituksellisen oloinen kansanomaisuus yhtäällä yhdistyy eleganttiin ja harkittuun korkeakirjalliseen tyyliin toisaalla. Mahdollisesti grammatikkamieltymyksensä takia Varro käyttää paljon arkaaisia ja suorasukaisia ilmauksia, mutta samaan aikaan tekstissä on nähty piirteitä koristeellisesta ja rönsyilevästä asianismista. Tarkempi tutkimus kuitenkin osoittaa tällaisen vastakkainasettelun liioitteluksi ja esimerkit aasisalaisesta tyylistä Varron kerronnassa jäävät todellisuudessa yksittäisiksi. Vastaavasti Varro näyttää myös useimmiten käyttävän täysin klassista latinaa, jollei esimerkiksi arkaismi ole asiayhteydessä perusteltu.

On kuitenkin ilmeistä, että Varrolle asiasisältö on (ainakin säilyneiden) tekstien pääasia, eikä hän ole pitänyt tarpeellisenä tai ensisijaisena käyttää aikaansa ilmaisun hiomiseen ja

⁵⁰ Heurgon 1978, XLV–L.

⁵¹ Laughton 1960, 1.

viimeistelyyn. Jo tuotannon valtavasta määrästä voidaan päätellä, että kirjoitusvauhdin on pakko olla ollut hyvin nopea, eikä lauseiden viimeistelyyn, saati korjailuun, ole jäänyt aikaa. On arveltu, ettei Varro ole välttämättä ehtinyt edes itse lukea kerran kirjoittamiaan tekstejä. Hyvin suuri osa pitkistä ja sekavista virkkeistä, anakoluuteista, oudosti sijoitetuista sivulauseista sekä kongruenssivirheistä (jotka tosin useimmiten ovat arkikielessä aivan tavanomaista *constructio ad sensum* -tyyppiä) selittyy juuri viimeistelemättömyydellä. Tekstin tarkemman analyysin kannalta onkin tarpeellista erottaa tällaiset ”huolimattomuusvirheet” rakenteista, jotka ovat Varron kielessä säännöllisiä ja tyypillisiä.

Viimeistelemättömyys

Vaikka huolimattomuusvirheet voivat tietenkin olla hyvin monenlaisia, Varron kielestä voidaan tunnistaa joitakin tyypillisiä ominaispiirteitä, jotka voidaan laskea viimeistelemättömyyden syyksi. Kaikkien piirteiden osalta ei ole aivan selvää, onko kyse viimeistelemättömyydestä vai tietoisemmasta, Varron kielenkäytölle ominaisesta ilmiöstä. Laughtonin jaottelusta poiketen esitän tässä esse-verbin poisjätön ja gerundiivin käyttötavan ominaispiirteenä pikemminkin kuin huolimattomuutena. Muita viimeistelemättömyyden piirteitä ovat pronomien ja jopa substantiivien poisjättäminen tapauksissa, joissa niitä normaalisti odotetaan, subjektin vaihto, kongruenssin häilyvyys, sekä anakoluutti. Tämän työn kannalta epäsäännönmukaisuudet voivat heikentää esimerkiksi luvussa 4.3 esitettyjen tietokoneavusteisten menetelmien kykyä tunnistaa lauserakenteita oikein.

Varro jättää usein pois myös pronomien ilmisubjektina – ja joskus myös objektina – sellaisissakin tapauksissa, joissa sen normaalisti odottaisi olevan näkyvillä. Kolmannen persoonan pronominit saattavat puuttua myös *accusativus cum infinitivo* -rakenteista. Varsinkin *Res Rustica*essa ilmisubjekti puuttuu yksikön kolmannessa persoonassa olevan predikaatin yhteydessä usein myös silloin, kun subjektiksi on ajateltavissa esimerkiksi joku tarkemmin yksilöimätön maanviljelijä. Esimerkiksi *diebus viginti antequam tollere vult turdos*.⁵² Useimmiten – joskaan ei aina – tällaisten tapausten läheisyydessä on kuitenkin gerundi-, gerundiivi- tai *oportet*-rakenne, joka tekee tarkoitetun subjektin ilmeiseksi. Toisinaan vastaavissa tapauksissa subjektina tai objektina oleva substantiivikin voi jäädä pois tai subjekti implisiittisesti vaihtua kesken virkkeen.

Vaikka Varro näyttää toisinaan lipsuvan tiukasta suku- ja persoonakongruenssista, nämä tapaukset tuskin kielivät perustavaa laatua olevasta kielioppisääntöjen sivuuttamisesta. Oikeastaan ainoa selkeä poikkeama sukukongruenssista on *res*-sanana mieltäminen neutriksi

⁵² *Rust.* 3,5,4.

etenkin monikossa, esimerkiksi *circumcisis rebus quae non arbitror*.⁵³ Tämä ei kuitenkaan ole mitenkään tavatonta muillakaan saman aikakauden kirjoittajilla. Lukukongruenssia Varro sen sijaan rikkoo ilmeisen usein ja vaihtaa yksiköstä monikkoon tai päinvastoin, vaikka viittaa koko ajan samaan sanaan.

Varsinkin puhekielessä on tavallista, että virkkeen pituuden kasvaessa ajatus katkeaa, ja virkkeen osat eivät sovi lauseopillisesti tai loogisesti yhteen. Tällaiset anakoluutit ovat Varron kielessä sangen yleisiä ja ilmeisen tahattomia. Erityisen helposti anakoluutti syntyy, jos virkkeen keskelle on upotettuna selittävä sivulause, mutta Varron tekstissä yhteensopimattomia virkkeen osia esiintyy rakenteeltaan yksinkertaisemmissakin tapauksissa.

Seuraavassa esimerkkikatkelmassa on useampia edellä mainituista piirteistä. Virke alkaa *oportet*-rakenteella, johon liittyy aivan säännönmukaisesti *accusativus cum infinitivo*. Vaikka varsinkin myöhäislatinassa myös *ut*-rakenne esiintyy *oportet*-verbin kanssa yleisesti, Varro tuskin olisi tällaista tarkoituksellisesti käyttänyt, ja virkkeessä myöhemmin tuleva *ut habeat* on mitä ilmeisimmin akoluutin seurausta. Kyseiseen kohtaan liittyy myös hieman jäljempänä käsiteltävä anastrofi. Seuraavassa virkkeessä subjekti, joksi on ajateltava edellisen virkkeen *venter*, muuttuu maskuliinista neutriksi, sillä siihen viitataan pronomiinilla *id*.

*de forma ovem esse oportet corpore amplo, quae lana multa sit et molli, villis altis et densis toto corpore, maxime circum cervicem et collum, ventrem quoque ut habeat pilosum. itaque quae id non habent, maiores nostri apicis appellabant ac reiciebant.*⁵⁴

Muodoltaan lampaan tulee olla isovartaloinen, jossa on paljon ja pehmeää villaa, koko vartaloltaan pitkä- ja tuuheaturkkinen, erityisen paljon niskan ympäristössä, myös että sillä on villainen vatsa. Niinpä niitä, joilla sitä ei ole, esi-isämme kutsuivat kaljuiksi ja hylkäsivät.

Varrolle tunnusomaiset rakenteet

Varro on mieltynyt käyttämään tiettyjä lauserakenteita ja sanajärjestyksiä, jotka tekevät tekstistä omintakeista verrattuna klassisen ajan muihin kirjailijoihin. Erityisen luonteenomaisia ja silmiinpistäviä piirteitä Varron kielessä ovat alistuskonjunktoiden, pronominien ja adjektiivien anastrofi, sivulauseen subjektin prolepsis sekä alisteisten sivulauseiden – etenkin relatiivilauseiden – sijoittaminen virkkeessä epätyypillisiin paikkoihin. Lisäksi etenkin *esse*-verbin poisjättöä sekä gerundi- ja gerundiivirakenteiden käyttötapoja voidaan pitää Varrolle luonteenomaisina.

Anastrofi tarkoittaa syntaktisen elementin – sanan tai lausekkeen – siirtämistä sen sanan jälkeen, jonka edellä se tavanomaisesti olisi. Varrolla anastrofi on erityisen tavallinen normaalisti sivulauseen aloittavien konjunktoiden, pronominien ja adverbien kohdalla.

⁵³ *Rust.* 1,1,11.

⁵⁴ *Rust.* 2,2,3.

Tällainen on varsin yleistä varhaisemmassa latinassa, mutta Cicerolla ja Caesarilla huomattavan harvinaista. Anastrofit voidaan jakaa kolmeen päätyyppiin: (A) siirretyn elementin edellä on yksi lauseke, (B) elementti on siirretty välittömästi verbin edelle ja (AB) verbin edelle siirrettyä elementtiä edeltää yksi lauseke.

Varrolla kaiken kaikkiaan tavallisinta on relatiivipronominien anastrofi, mutta *Res Rustica*essa *ut-* ja *ne-*konjunktioiden anastrofi on vähintään yhtä yleistä. Muiden konjunktioiden, interrogatiivipronomininen ja adverbien kohdalla ilmiö on selvästi harvinaisempi. *Res Rustica*essa yleisin anastrofi on välittömästi verbin edelle siirtynyt elementti (B-tyyppi) ja A-tyyppi harvinaisin. Edellä olevassa esimerkxivirkkeessä on B-tyyppinen *ut*-konjunktion anastrofi *ventrem quoque ut habeat pilosum*. Esimerkki A-tyypin anastrofista, jossa siirretty elementti on useamman sana lauseke:

*et repentinae praedonum manus **quod** improvisos facilius opprimere possunt*⁵⁵

AB-tyyppinen anastrofi taas on esimerkiksi:

*meliore genere **ut** SIT surculus, quam est quo veniat arbor*⁵⁶

Varron anastrofeille on myös tunnusomaista, että siirretty syntaktinen elementti voi olla kokonainen lause, esimerkiksi:

*siquem glabrum facere velis **quod** IUBET ranam luridam coiecere in aquam*⁵⁷

Anastrofeja tässä työssä tutkittavassa aineistossa tarkastellaan vielä esimerkinomaisesti puupankkianalyysin yhteydessä luvussa 6.3.

Samantapainen ilmiö kuin anastrofi on prolepsis, jossa alisteisen sivulauseen subjekti on siirretty lauseen edelle, päälauseen predikaatin objektiksi. Kuten anastrofikin, prolepsis on yleinen varhaisessa latinassa, ja ilmeisesti myöhemminkin puhekielessä. Varrolla tästä on useita esimerkkejä, joskaan ei yhtä paljon kuin anastrofista, ja tapaukset esiintyvät usein yhdessä muiden lauserakennetta hämärtävien piirteiden, kuten anastrofin tai erikoisesti sijoitetun relatiivilauseen, kanssa.

Relatiivilauseiden tavallisuudesta poikkeava käsittely onkin kolmas Varron kielen rakenteille luoteenomainen seikka. Erityisen silmiinpistävää on relatiivilauseen sijoittaminen ennen korrelaattiaan, korrelaatin attraktio relatiivilauseeseen ja relatiivilauseiden mielivaltaiselta näytävä sijoittelu virkkeeseen. Kun korrelaatti ja relatiivipronomini ovat samassa sijassa, relatiivilause ennen korrelaattia on melko tavallinen ilmiö yleensäkin latinankielisessä kirjallisuudessa. Varrolla tämä on kuitenkin erityisen yleistä ja lisäksi

⁵⁵ *Rust.* 1,12,4.

⁵⁶ *Rust.* 1,40,6.

⁵⁷ *Rust.* 1,2,26.

pronomini ja korrelaatti voivat olla myös eri sijoissa. Vastaavasti samassa sijassa korrelaattina oleva pronomini jätetään latinassa usein pois, mutta Varro ulottaa tällaisen omission myös eri sijassa olevaan korrelaatiin. Lisäksi Varrolla on tapana lisätä relatiivilauseita erikoisiin paikkoihin virkkeessä ilman ilmeistä perustetta, ikään kuin suunnittelemattomina lisäyksinä. Muunlaisten sivulauseiden sijoittelu sitä vastoin ei määrällisesti näytä Varrolla juuri eroavan muista aikakauden auktoreista, mutta niiden siirtäminen ennen korrelaattia on usein silmiinpistävää, sillä ne tapaavat liittyä muihin sanajärjestyksen poikkeavuuksiin, kuten anastrofeihin.

Esse-verbin indikatiivin presensin kolmannen persoonan muodot *est* ja *sunt* jäävät Varron tekstissä hyvin tavallisesti pois sekä kopulana että partisiipin, gerundin ja gerundiivin yhteydessä apuverbinä. Myös infinitiivimuoto *esse* jää usein pois varsinkin sanomista ja ajattelemista ilmaisevien verbien kanssa. Muut persoona- ja aikamuodot sekä konjunkttiivissa oleva kopula sen sijaan ovat tavallisesti näkyvillä. *De Lingua Latina* -teoksessa myös muut verbit kuin *esse* jäävät melko yleisesti pois, mutta *Res Rusticae* muistuttaa tässä mielessä vähemmän muistiinpanoja eikä verbien poisjätto ole tavallista.

Gerundiivin käytössä kokonaisuudessaan näkyvät Varron tyylille ominaiset arkaaiset ja puhekieliset ilmaukset sekä siellä täällä kiireen aiheuttamat epäsäännöllisyydet. Aikalaistensa normaalikäytännöstä poikkeava usein, joskaan ei säännönmukaisesti, toistuva piirre on gerundiivin käyttäminen persoonattomasti suoran akkusatiiviobjektin kanssa, esimerkiksi *villam aedificandum*.⁵⁸ Toisinaan gerundi- tai gerundiivilausekkeeseen voi myös liittyä akkusatiivissa oleva partisiipin presens, joka viittaa johonkuhun tekstissä tarkemmin yksilöimättömään, Varron mielessä olevaan henkilöön, esim. *queadam etiam serenda non tam propter praesentem fructum quam in annum prospicientem*.⁵⁹

⁵⁸ *Rust.* 1,11,2.

⁵⁹ *Rust.* 1,23,3.

4 AINEISTO JA TEKSTIN PROSESSOINTI

Tämän työn aineistona on *Res Rusticae*n kaikki kolme kirjaa käsittävä tekstikorpus. Tässä luvussa tarkastelen kyseisen tekstin käsikirjoitustraditiota ja rekonstruktiota sekä tekstin digitaalisen version prosessointia. Jälkimmäisen osalta esittelen tekstin käsittelyssä noudattamiani periaatteita, käyttämiäni tietokoneavusteisia menetelmiä sekä näiden avulla myöhemmän analyysin perustaksi luotuja tekstietokantoja. Varsinaista tekstianalyysiä ja tietoteknisten välineiden taustalla olevia kielitieteellisiä menetelmiä käsittelen tarkemmin metodologian yhteydessä luvussa 5.

4.1 Käsikirjoitukset ja tekstin rekonstruktio

Keskiaikaisia käsikirjoituksia, jotka sisältävät *Res Rusticae*n kokonaan tai osittain, on säilynyt yhteensä 34 kappaletta. Käsikirjoitukset on jaettu stemmatologiassa viimeaikoihin asti kahteen päähaaraan, joiden yhteistä arkkityyppiä pidetään kadonneena.⁶⁰ Uusin tutkimus on kuitenkin päätymässä johtopäätökseen, että kaikille tunnetuille käsikirjoituksille yhteinen arkkityyppi on itse asiassa codex *Marcianus* [Ω] -nimellä tunnettu, nyttemmin myös kadonnut teksti.⁶¹ Jo tekstistä ensimmäisen varsinaisen kriittisen edition laatinut Heinrich Keil oli pitänyt *Marcianusta* kaikkien säilyneiden käsikirjoitusten arkkityyppinä. Sama käsikirjoitus sisälsi myös Caton *De agricultura* -teoksen, jonka varhainen käsikirjoitukseen pohjautuva tekstihistoria on *Res Rusticae*n kanssa yhtenevä. Codex *Marcianus* oli ilmeisesti laadittu karolingisen renessanssin aikana 800- tai 900-luvulla, ja sitä oli säilytetty Firenzessä, pyhän Markuksen luostarin kirjastossa.

Itse käsikirjoitus on kadonnut, mutta lukutavat tunnetaan, sillä Angelo Poliziano teki vuonna 1482 sen pohjalta korjaukset Giorgio Merulan 1472 julkaisemaan *editio princeps*iin; tämä korjausmerkinnät sisältävä kappale (Ω^P) on nykyisin Ranskan kansalliskirjastossa Pariisissa. Vaikka Poliziano jättikin joitakin Merulan edition virheitä huomiotta, Ω^P:tä voidaan kuitenkin pitää [Ω]:n jotakuinkin täydellisenä toisintona. Lisäksi Ω^P oli Piero Vettori (1499–1585) käytössä, kun hän laati vuonna 1541 julkaisemaansa editiota Ω^V. Sekä Poliziano että Vettori ovat kopioineet Caton 84. luvun kohdalla marginaali huomautuksen, jonka mukaan *Marcianuksen* kirjurilla oli siihen asti ollut käytössään kaksi eri käsikirjoitusta, mutta tämän kohdan jälkeen vain yksi. Näin ollen käsikirjoituksessa Caton jälkeen tuleva *Res rusticae* perustuu jo tässä vaiheessa yhteen ainoaan lähteeseen.

⁶⁰ Esim. Flach 2006, Heurgon 1978.

⁶¹ Rodgers 2015:161. Stemman kirjainmerkinnät seuraavat tässä Rodgersin esitystä.

Säilyneistä käsikirjoituksista vanhin ja parhaana pidetty codex *Parisinus Latinus* 6842 (A) on ollut lähtökohtana sukupuun toiselle haaralle. Se myös sisältää kolmannen kirjan kokonaisuudessaan ilman Marcianuksessa kohdassa 3.7.4–10 olutta lacunaa. Käsikirjoitus on 65-sivuinen, kirjoitettu karolingisella minuskelilla 1100- ja 1200-lukujen taitteessa ja siinä on myös eri käsialalla tehtyjä korjausmerkintöjä. Se on mahdollisesti kuulunut Petraralle (1304–74), jonka kirjoista suuri osa päättyi myöhemmin Paviassa Visconti-Sforzien kirjastoon, jossa tiedetään vuonna 1426 olleen kaksi *Res Rusticae*n sisältänyttä käsikirjoitusta. Näistä yksi on kadonnut ja toinen identifioitavissa A:ksi.

Codex *Laurentianus* 51,1 (F) on peräisin 1300-luvulta ja kirjoitettu goottilaisella minuskelilla. Se luetaan A-perheeseen, mutta sitä on korjattu jonkin toisen varhaisen käsikirjoituksen perusteella. A-perheeseen kuuluva joukko muita käsikirjoituksia 1400-luvulta – *Laurentianus* 51,3, *Parisinus Latinus* 11213, *Laurentianus* 30,10 ja *Neapolitanus* V A 8 – polveutuu eri reittejä A:n lisäksi F:stä. Nämä *recentiores*-käsikirjoitukset ovat pääosin huolimattomasti kopioituja tai humanistien yli-innokkaasti korjaamia, mutta niistä oli kuitenkin merkittävää apua humanisteille, jotka eivät tunteneet A:ta.

Suoraan *Marcianuksesta*, mukaan lukien kolmannen kirjan lacuna, polveutuu codex *Laurentianus* 51,4 (B), joka on peräisin 1400-luvulta kuten kaikki muutkin *recentiores*-käsikirjoitukset. Sen pohjalta voidaan vahvistaa useimmat Ω^P :n lukutavat. Muista *recentiores*-käsikirjoituksista *Laurentianus* 51,2, *Vindobonensis* 33 H ja *Malatestianus* S 24,2 ovat olleet jossakin määrin ediittorien käytössä. Merulan *editio princeps* – *Scriptores rei rusticae* joka sisältää Varron lisäksi Caton, Columellan, ja Palladiuksen tekstit – on laadittu Varron ja Caton osalta käsikirjoitusten F, B, *Laurentianus* 30,10 ja *Laurentianus* 51,3 pohjalta.

*

*Res Rusticae*n ensimmäinen, Keilin kokoama varsinainen kriittinen editio on vuodelta 1884 (*editio maior*). Vuonna 1889 Keil julkaisi tästä korjatun version (*editio minor*) ja näiden pohjalta hänen seuraajansa Georg Goetz Teubner-editiot 1912 ja 1929. Myös Loeb-editio vuodelta 1934 ja Budé-editio⁶² (1978-1997) seuraavat hyvin tarkasti Keilin ja Goetzin tekstiä. Nämä editiot ovat pitäneet *Marcianuksesta* lähtöisin olevia lukutapoja pääsääntöisesti luotettavimpina. Ainakin viidessä ensimmäisen kirjan kohdassa Heurgon kuitenkin pitää A:n antamaa lukutapaa selvästi oikeampana. Lisäksi sekundäärilähteet ovat useissa tapauksissa hyödyllisiä: Heurgonin mielestä Noniuksen sitaateissa (ks. edellä luvut 2.2 ja 2.3) annettu

⁶² Heurgon 1978, Giuraud 1985, 1997.

lukutapa on oikea (joko vahvistaen käsikirjoituksen lukutavan tai muuttaen sitä) 13 tapauksessa 21:stä.

Toistaiseksi uusimman edition on julkaissut Dieter Falch vuosina 1996-2002, mutta tätä laitosta pidetään ongelmallisena varsinkin tekstikritiikin osalta. Näiden lisäksi on vuonna 2006 julkaistu Michelangelo Giustan keskeneräiseksi jääneen edition *Res Rusticae* kaksi ensimmäistä kirjaa. Roger H. Rodgers, joka on laatimassa kokonaan uutta kriittistä editiota, huomauttaa edeltäjiensä pitkälti takertuneen käsikirjoitustraditioon – mikä kuitenkin on suureksi osaksi virheiden vertailua keskenään – koska Varron tekstin viimeistelemättömyys sekä luultavasti tarkoitukselliset arkaaiset, arkikieliset ja substandardit ilmaisut tekevät vaikeaksi erottaa alkuperäistekstin ominaispiirteitä kirjurien tekemistä kopiointivirheistä tai humanistien liian innokkaasta korjauksista.⁶³ Varro itsekin toteaa koonneensa teoksen kovalla kiireellä julkaisua varten, mikä näkyy tekstissä monin paikoin viimeistelemättömyytinä.

Olen tässä käyttänyt epäselvien kohtien tulkinnassa ensisijaisesti Budé-edition lukutapoja, mutta tekstin saatavuuden takia digitaalinen aineisto pohjautuu Loeb-edition tekstiin pienin korjauksin, jotka on esitetty seuraavassa luvussa. Tämän työn kannalta merkittävin haaste käytettävissä olevissa editioissa on puhujien paikoin puutteellinen tai virheellinen identifiointi. Esimerkiksi Rogers mainitsee kohdan 1,1,2 (mahdollisesti puhujana välillä Varro itse)⁶⁴ ja Thayers kohdan 2,11,10 (sisällöstä päätellen puhuja näyttää vaihtuneen).⁶⁵ Heurgon mainitsee vastaavia epäselvyyksiä kohdissa 1,8,1 ja 1,15.⁶⁶

4.2 Digitaalinen aineisto

Digitaalisen tekstianalyysin aineistona on *Res Rusticae* kaikki kolme kirjaa. Tekstikorpus on vapaasti saatavilla useammasta lähteestä. Tässä työssä pohjana olen käyttänyt Lacus Curtius -sivustolla⁶⁷ olevaa Loeb Classical Libraryn digitoitua tekstiä⁶⁸, joka puolestaan pohjautuu Gaezin Teubner-edition, kuten edellä on mainittu. Olen muokannut tätä raakatekstiä paremmin tietokoneella automaattisesti prosessoitavaksi alla tarkemmin kuvatuilla tavoilla. Koska automaattisessa käsittelyssä on käytetty useita eri työkaluja, jotka osittain vaativat syötteen erilaisissa muodoissa ja raportoivat tulokset erilaisten standardien mukaisesti, olen

⁶³ Rodgers 2015.

⁶⁴ Rodgers 2015, 170.

⁶⁵ Thayers,

https://penelope.uchicago.edu/Thayer/E/Roman/Texts/Varro/de_Re_Rustica/2*.html#note:digression_on_barber_S

⁶⁶ Heurgon 1978, XLV–L.

⁶⁷ Thyer; http://penelope.uchicago.edu/Thayer/E/Roman/Texts/Varro/de_Re_Rustica/home.html

⁶⁸ Hooper & Ash 1934.

koonnut kaikki käytetyt versiot ja tulokset tietokannaksi. Tietokannan avulla eri tuloksia voidaan vertailla keskenään ja tietoja yhdistellä keskenään. Kaikkiaan aineistossa ilman välimerkkejä noin 35 600 sanetta, välimerkit mukaan lukien noin 42 700 sanetta.

Replikkien erottelu ja puhujien tunnistus

Koska tarkoitukseni on tarkastella dialogimuotoisen tekstin mahdollisia kielellisiä eroavaisuuksia eri puhujien välillä, repliikit on tunnistettava tekstistä. Tätä ei ole ainakaan sanatasolla nähtävästi aikaisemmin tehty. Hieman karkeammalla tasolla esimerkiksi Nelsestuen on kyllä maininnut myös puhujat Varron lähteiden tarkastelun yhteydessä.⁶⁹ Tätä työtä varten olen tunnistanut puhujat virkkeittäin lukemalla koko tekstin kursorisesti läpi. Vaikka tämä on pääosin melko suoraviivaista, tunnistamiseen liittyy kuitenkin ongelmia, kuten edellä käsikirjoitustradition yhteydessä mainittiin. Puhujat olen merkinnyt erilliseen taulukkoon, joka sisältää erilaisia metatietoja virketasolla ja josta ne voidaan yhdistää kaikkiin analysoitaviin saneisiin (tietokannan tarkempi esitys jäljempänä). Tämän alustavan puhujien tunnistuksen tulokset tarkastetaan ja tarvittaessa korjataan vielä luvussa 5.2 esiteltävän tekstin kielellisen annotoinnin yhteydessä. Käytännössä puhujat esitetään annotoinnissa omana *speaker*-leimanaan.

Esikäsitteily, virkejako, saneistus ja annotointi

Jotta digitaalinen aineisto olisi hyödynnettävissä mahdollisimman tehokkaasti erilaisilla tietoteknisillä työkaluilla, tekstikorpusta on jonkin verran muokattava. Vaikka jotkin sovellutukset voivatkin käyttää tekstiä ilman muunnoksia, eri työkalujen tuottamien tietojen yhdistettävyyttä edellyttää alkuperäiseltä tekstitiedostolta sopivaa rakennetta. Kaikki alkuperäisessä digitaalisessa tekstissä olevat ns. ascii-lainausmerkit (") on korvattu kulmalainausmerkeillä (»), sillä ascii-lainausmerkit aiheuttavat lukuisia erilaisia virheitä aineiston koneellisessa käsittelyssä ja tavalliset lainausmerkit (") saattavat esimerkiksi tekstinkäsittelyohjelmissa muuttua kieliversion mukaan käyttäjän huomaamatta. Lisäksi kaikki välimerkit on erotettu prosessoinnin ensimmäisessä vaiheessa välilyönnillä edeltävästä sanasta, koska jotkin sovellukset käsittelevät välimerkin sisältäviä merkkijonoja arvaamattomasti.

Tekstin jakaminen virkkeisiin ei ole kielellisen analyysin kannalta triviaalia, ja automaattisesti suoritettu jako saattaa vaikuttaa lopputulokseen mielivaltaisesti. Tietotekniset sovellukset tavallisesti tulkitsevat virkerajaksi pisteen, kysymysmerkin, kaksoispisteen ja puolipisteen. Jos pisteen muuta käyttöä (esim. lyhenteet) ei tunnisteta oikein, virke voi katketa

⁶⁹ Nelsestuen 2015, 246–61.

väärästä kohdasta. Puolipisteen käytössä on ediittoreiden välillä eroja, jotka osin johtunevat eri kielten käytännöistä. Johtolauseita sisältävässä dialogiteksteissä kaksoispisteet ja tietyissä tapauksissa myös kysymysmerkit aiheuttavat erikoisia tilanteita. Esimerkiksi käy seuraava lause Budé-edition (Rust 1.2.3)

Cum consedissemus, Agrasius: Vos, qui multas perambulastis terras, ecquam cultiorem Italia vidistis? inquit.

Kun olimme istuutuneet, Agrius sanoi: Te, jotka olette kolunneet monia maita, oletteko nähneet yhtään Italiaa viljavampaa?

Jos pitäydymme automaattisesti tuotetussa jaossa, joka pitää kaksoispistettä ja kysymysmerkkiä virkkeen lopun merkinä, tämä esimerkki jakautuisi kolmeen eri virkkeeseen. Näistä ensimmäinen, *Cum consedissemus, Agrasius*, ja viimeinen, *inquit*, jäisivät oudosti erilleen. Lisäksi virkejako muuttuu mielivaltaisesti sen mukaan, missä kohdin virkettä johtolause on. Saatavilla olevissa teksteissä on käytetty isoa alkukirjainta virkkeiden alussa, paitsi puolipisteen jälkeen, joten suuri osa lyhenteistä ja kysymysmerkkiä seuraavista johtolauseista voidaan tunnistaa automaattisesti tällä perusteella. Kaksoispistettä en ole normaalikäytännöstä poiketen pitänyt tässä työssä automaattisesti virkkeen lopun merkinä, vaan nämä jätetään arvioitaviksi tapauskohtaisesti manuaalisen annotoinnin yhteydessä.

Latinankielisten tekstien sanastolliset analyysit, kuten frekvenssilaskelmat, ovat mielekkäitä vain, jos saneista (engl. tokens) on käytettävissä myös hakusanamuodot (lemmat). Lemmojen tunnistukseen liittyy myös saneistus, jossa tekstin merkkijonot jaetaan sanakirjahakua tai morfologista analyysia varten soveltuviksi saneiksi, jotka on normalisoitu ja puhdistettu tai erotettu välimerkeistä.⁷⁰ Saneistukseen on saatavilla sekä geneerisiä että latinaa varten räätälöityjä sovelluksia. Geneeriset mallit eivät tunnista esimerkiksi latinan liitepartikkeleja ja räätälöidyissäkin sovelluksissa tulokset ovat vaihtelevia. Kokeilujen perusteella paras saneistustulos syntyy Arethusa-työkalulla⁷¹, jonka tuottamaa xml-tiedosta olen käyttänyt jatkokäsittelyn pohjana.

Koko virkkeisiin jaettu ja saneistettu, xml-muotoinen tekstikorpus on muunnettu Conference on Computational Natural Language Learning CoNLL-U -standardin mukaiseen muotoon.⁷² Tässä muodossa jokaiseen saneeseen voidaan liittää – annotoida – erilaisia tietoja, kuten lemmat, morfologiset piirteet ja puhuja. Manuaalisen annotoinnin yhteydessä voidaan

⁷⁰ Tieteen termipankki, 1.8.2020: Nimitys:saneistaa. (Tarkka osoite: <https://tieteen termipankki.fi/wiki/Nimitys:saneistaa>.);

⁷¹ <https://www.perseids.org/tools/arethusa/app/#/>

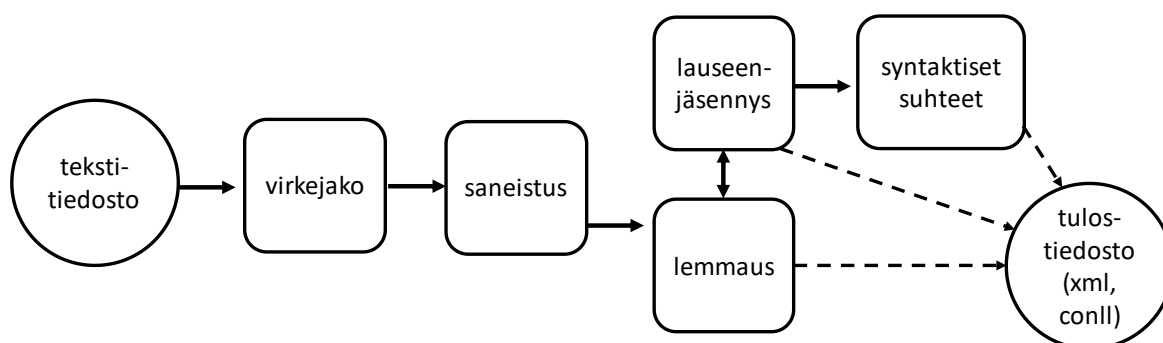
⁷² Buchholz & Marsi 2006; <https://universaldependencies.org/format.html>

myös tarkastaa ja korjata edellä mainitut epäselvät tai tulkinnanvaraiset kohdat. Lemmausta ja kielellisten piirteiden annotointia käsitellään metodologiselta kannalta luvussa 5.

4.3 Tietotekniset apuvälineet ja tietokanta

Jotta analysoitava aineisto olisi mahdollisimman laaja, olen käyttänyt hyväksi tietokoneavusteisia menetelmiä, erityisesti automaattista saneistusta, lemmausta ja jäsennystä (eng. parsing). Näitä varten on saatavilla useita työkaluja, jotka käyttävät ja tuottavat tekstiä eri muodoissa. Tämän työn kannalta keskeisimmät tiedostomuodot ovat puhdas teksti (*plain text*), edellä mainittu Conll-U sekä Text Encoding Initiative (TEI) -standardin mukainen Extended Markup Language (XML). Tekstin muunnokset, analyysi ja annotointi voidaan kuvata etenemiskaaviona, jossa eri vaiheet kuvataan niiden tyypillisessä toteuttamisjärjestyksessä. Kuviossa 1 on esitetty yksinkertaistettu etenemiskaavio tässä työssä käytetyn aineiston käsittelystä.⁷³

Kuvio 1: Aineiston käsittelyn eteneminen



Lähtökohtana olevasta tekstitiedostosta on ensin eroteltu virkkeet edellä mainituilla periaatteilla. Tässä vaiheessa pääasiallisena tietoteknisenä välineenä on Notepad++ -tekstieditori⁷⁴, joka soveltuu hyvin monenlaisten tekstiaineistojen käsittelyyn ja tukee mm. säännöllisiä lausekkeita (*regular expression*). Lisäksi omille riveilleen jaetut virkkeet tallennetaan taulukkomuodossa, jolloin niitä voidaan käsitellä taulukkolaskenta- ja tilasto-ohjelmilla (Excel, Stata). Tämä helpottaa eri versioiden vertailua, sillä virkkeet – ja myöhemmässä vaiheessa myös saneet – voidaan yksilöidä rivinumerotunnistein, joiden avulla esimerkiksi välimerkeistä puhdistetun tekstitiedoston saneet voidaan kohdistaa alkuperäiseen versioon.

⁷³ Etenemiskaavion tyypillinen rakenne ja käyttö antiikin kielten tutkimuksessa, ks. Burns (2019): 148–151.

⁷⁴ <https://notepad-plus-plus.org>

Edellä mainittu Arethusa on annotointityökalu, joka on kehitetty ensisijaisesti Ancient Greek and Latin Treebank -standardien⁷⁵ muotoisten puupankkien käsittelyyn. Työkalun käyttöliittymä on helppokäyttöinen ja siihen integroitu Morpheus-työkalu helpottaa morfologista analyysiä. Automaattisen virkejaon ja saneistuksen lopputulos on myös hyvälaatuinen. Vaikka olen tässä työssä päätenyt käyttämään pääosin Universal Dependencies -standardin mukaista annotointia, jonka käyttöä ei toistaiseksi tueta Arthusassa⁷⁶, työkalun tuottama virkejako ja saneistus koko tekstikorpukselle ovat kuitenkin pohjana jatkokäsittelylle. Lisäksi olen annotoinut Arethusalla ALDT-standardin mukaisesti 970 saneen katkelman, jota käytetään UD-aineiston vertailuaineistona luvussa 5.2. Arethusan tuottamat xml-muotoiset tiedostot on muunnettu Excelin ja Notepad++:n avulla CoNLL-U-muotoon, ja tuloksena olevaan tiedostoon yhdistetään edellä mainitut metatiedot virkkeittäin niin, että kullekin saneelle lisätään mm. puhujan osoittava ”Speaker”-leima.

CoNLL-U-muotoisia puupankkitiedostoja voidaan hyödyntää useissa sovelluksissa, kuten web-pohjaisessa Tundra-analyysityökalussa⁷⁷ ja jäljempänä käsiteltävässä UDPipe-jäsentimessä. Näiden tiedostojen editointiin ja annotointiin on saatavilla useita työkaluja, joiden ominaisuudet vaihtelevat. Usein näiden sovellusten asennus ja joskus myös käyttö vaatii ainakin pintapuolista ohjelmointikielten osaamista, mutta käyttöliittymät ovat kuitenkin yleensä selainpohjaisia. Tässä työssä pääasiallisena annotointi- ja editointityökaluna on käytetty Java-pohjaista ConlluEditor-sovellusta.⁷⁸ Käyttöliittymä on selkeä mutta monipuolinen, ja annotointileimojen lisäksi työkalulla voi tarvittaessa helposti muuttaa sane- ja virkejakoa, mikä useimmissa muissa sovelluksissa edellyttää lähdetiedoston editointia suoraan. Morfologisen analyysin apuvälineitä ei tosin saa yhtä saumattomasti käytettyä kuin Arethusa-sovelluksessa, mutta toisaalta integrointi UDPipe-jäsentimeen on mahdollinen.⁷⁹

Automaattinen jäsennin on varsinaisesti tietotekninen sovellus, joka analysoi tekstiä kielellisesti, esimerkiksi muoto- ja lauseopillisesti. Tässä työssä käytetty UDPipe⁸⁰ on itse asiassa prosessityökalu (pipeline), joka suorittaa virkejaon, saneistuksen, lemmauksen ja morfologisen analyysin, lauseenjäsennyksen ja dependenssisuhteiden analyysin

⁷⁵ annotointikehikoita käsitellään tarkemmin luvussa 5.2.

⁷⁶ Arethusaan on mahdollista luoda ja lisätä räätälöityjä annotointimalleja, ja esim. englanninkielisen tekstin Universal Dependencies tagset on käytettävissä. Tämän työn puitteissa työkalun räätälöinti ei kuitenkaan ole tarkoituksenmukaista.

⁷⁷ Martens 2013; <https://weblicht.sfs.uni-tuebingen.de/Tundra/about>

⁷⁸ Heinecke 2019.

⁷⁹ Manuaalin mukaan; ominaisuutta ei tässä työssä ole testattu.

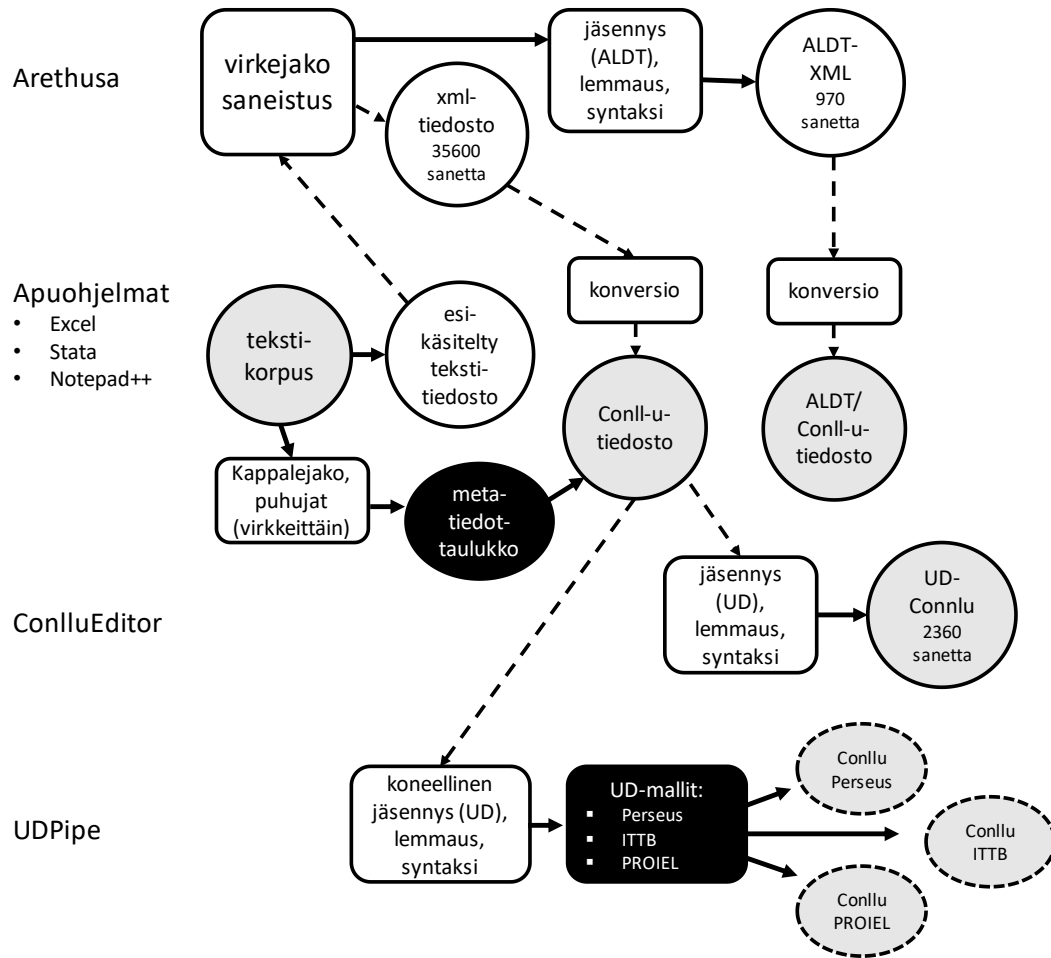
⁸⁰ Straka et al. 2016; Straka & Straková 2017; <https://lindat.mff.cuni.cz/services/udpipe/>

tekstistyyönteelle, kuten kuvion 1 etenemiskaaviossa on havainnollistettu. Dependenssisuhteiden mukaan järjestetyt saneet muodostavat puurakenteen, jota kuvataan tarkemmin jäljempänä luvussa 5.2. UDPipe on kieliriippumaton ja ”koulutettava”, eli sitä voidaan opettaa jäsentämään tietyn kielistä ja tyyppistä tekstiä antamalla malliksi valmiiksi analysoitua tekstiä CoNNL-U-puupankkina. Latinankieliselle aineistolle on saatavilla kolme mallia, jotka perustuvat eri puupankkeihin (Perseus, ITTB, PROIEL)⁸¹. Tässä tutkittavana olevalle tekstille on tuotettu automaattisesti muodostetut puupankit kaikilla kolmella mallilla. Työkalun virkejako- ja saneistusominaisuuksia ei kuitenkaan ole tässä hyödynnetty, vaan kuten edellä on mainittu, teksti on jo aiemmin jaettu virkkeisiin ja saneistettu Arethusalla. Tämä johtuu mm. siitä, että saatavilla olevat UD-mallit eivät tunnista saneita ja lauserajoja kovin hyvin – esimerkiksi PROIEL-malli perustuu korpuksen, josta on poistettu välimerkit. Muutenkin automaattisen jäsennyksen tulokset ovat vaihtelevia, ja tuloksia onkin hyödynnetty pääasiassa lemmafrekvenssianalyysissä.

Kuviossa 2 on koottu yhteen eri työvaiheet (laatikot), niissä käytetyt tietotekniset työkalut (vasemman reunan sarake) sekä tiedostot (ympyrät), joita prosessissa on käytetty ja syntynyt. Tuloksena olevia CoNNL-U-tiedostoja voidaan tarkastella esimerkiksi ConlluEditorilla ja hyödyntää haku- ja analyysityökaluissa, kuten aiemmin mainitussa Tundra-sovelluksessa.

⁸¹ Myös näitä käsitellään lähemmin seuraavassa luvussa.

Kuvio 2: työvaiheet, tietotekniset työkalut, tiedostot



5 METODOLOGIA

Tässä luvussa käsittelen analyysissä käytettyjä kielitieteellisiä menetelmiä. Olen päätenyt käyttämään dialogin henkilöiden kielen eroavaisuuksien tunnistamiseen kahta metodia, jotka mahdollistavat tekstikorpuksen tilastollisen analyysin. Tekstissä esiintyvien sanojen ja erityisesti niiden lemmojen esiintymismäärien vertailu, frekvenssianalyysi, antaa karkean mutta suhteellisen kattavan kuvan mahdollisista leksikaalisista eroista. Erityisesti funktiosanojen frekvenssien tarkastelu voi kertoa myös eri puhujien kielessä esiintyvistä rakenteellisista piirteistä. Syntaktinen annotointi taas on kunnolla tehtynä työläs menetelmä, mutta lopputuloksena antaa mahdollisuuden tutkia tekstin piirteitä hyvin monipuolisesti. Menetelmän tekee vaivan arvoiseksi myös se, että korpusten analyysissä yleisesti käytettävät sanaparien etäisyyksiin perustuvat menetelmät soveltuvat huonosti kirjalliseen latinaan, jossa toisiinsa liittyvät sanat saattavat sijaita monimutkaisissa virkerakennelmissa hyvinkin kaukana toisistaan.

5.1 *Frekvenssianalyysi*

Kielellisten piirteiden esiintymismäärien tarkastelun pitäminen varsinaisesti metodina on ehkä hieman liioiteltua: Lähes kaikki korpustutkimus perustuu vähintään implisiittisesti sen arvioimiseen, onko jokin tietty sana, rakenne, ilmaisu, tyylikeino tms. jossakin yhteydessä tyypillinen tai yleinen joko itsesään tai verrattuna johonkin muuhun tekstiin. Toisin sanoen analyysissä tarkastellaan kielellisten ilmiöiden frekvenssejä. Varsinaisesti frekvenssianalyysillä kuitenkin viitataan korpuksen tilastollisia menetelmiä hyödyntävään tutkimukseen.

Tässä työssä käytetystä korpuksesta tilastollisen analyysin tekeminen on teknisesti suoraviivaista. Luvussa 4.3 kuvattu aineisto sisältää koko tekstin eroteltuna saneiksi, joista on jo tietokoneavusteisesti määritetty lemmat, taivutusmuodot ja sanaluokat. Lisäksi metatietoihin on merkitty dialogin puhujat, jolloin frekvenssejä puhujien välillä voidaan vertailla. Tässä tietokoneen automaattisesti tekemässä analyysissä on kuitenkin useita ongelmia. Ilmeisin haaste syntyy homonymiasta ja polysemiasta, joiden seurauksena joidenkin sanojen tietyt muodot lankeavat melkoisen sattumanvaraisesti eri hakusanoille (lemmoille). Toki homonymiaa on myös useiden taivutusmuotojen välillä, jolloin morfologinen analyysi saattaa olla virheellinen, vaikka lemma olisikin oikein.

Koska esiintymisfrekvenssien osalta tarkasteluni keskittyy funktiosanoihin, edellä mainitut ongelmat ovat jonkin verran lievempiä. Vaikka homonymia on edelleen mahdollista (esim. *cum* konjunktiona tai prepositiona, *uti* verbimuotona tai *ut*-konjunktion varianttina),

virhemahdollisuudet ovat kuitenkin huomattavasti rajatumpia ja helpommin tarkastettavissa eikä esimerkiksi eri taivutusmuotojen tunnistamisesta muodostu ongelmaa.

Johtopäätösten kannalta olennaista on tunnistaa, eroaako jonkin kielellisen piirteen esiintyminen puhujien välillä tilastollisesti merkitsevästi. Tarkastelun perusyksikkönä on esiintymien suhteellinen frekvenssi, eli esimerkiksi tietyn saneen esiintymien yhteenlaskettu määrä jaettuna saneiden kokonaismäärällä – kullakin puhujalla erikseen, kun tarkastellaan puhujien välisiä eroja. Näitä verrataan saman piirteen suhteelliseen esiintymisfrekvenssiin koko aineistossa. Tilastollisen merkitsevyyden arviointiin voidaan käyttää sopivia tilastollisia testisuureita tai tarkastella frekvenssien ohella niiden keskihajontoja.

5.2 Dependenssikielioppi ja korpusten kielellinen annotointi

Käytän tässä työssä tutkittavan korpuksen lauseopillisen analyysin tukena dependenssikieliopin mukaista syntaksin annotointia. Dependenssi- eli riippuvuuskieliopissa tarkastellaan lauseenosien välisiä syntaktisia riippuvuussuhteita. Erityisesti pyritään löytämään kustakin rakenteesta pääsana (regentti), jolle määritteet (dependentit) ovat alisteisia. Tältä osin dependenssirakenne eroaa puupankeissa myös yleisesti käytetystä konstituenttianalyysistä, jossa pääpaino on virkkeiden itsenäisesti käyttäytyvien osakokonaisuuksien määrittelemisessä.⁸² Korpustutkimuksen yhteydessä annotoinnilla tarkoitetaan tekstin varustamista kieliopillista rakennetta tai esimerkiksi saneiden ominaisuuksia ja kategorioita osoittavilla leimoilla (engl. tag) . Annotointi tehdään joko käsin tai (osittain) automaattisesti käyttämällä tietokoneavusteisia apuvälineitä, kuten jäsentimiä (engl. parser).⁸³

Dependenssikieliopin mukaisessa kielellisessä annotoinnissa yleisimmin käytetyt skeemat perustuvat joko Prague Dependency Treebank⁸⁴ (PDT) tai Universal Dependencies (UD) -kehikkoon. Ensimmäiset ja suurimmat antiikin tekstien syntaktisesti annotoidut korpuksat perustuvat kaikki PDT-kehikkoon. Ancient Languages Dependency Treebank⁸⁵ sisältää kreikan- (AGDT) ja latinankielisiä (ALDT), lähinnä klassillisen ajan tekstejä (50 000 latinankielistä sanetta). Index Thomisticus Treebank⁸⁶ (ITTB) käsittää Tuomas Akvinolaisen tuotantoa ja tähän läheisesti liittyviä muita keskiajan latinankielisiä tekstejä (450 000 sanetta).

⁸² Karlsson 2006:126–7, 156–7; Tieteen termipankki 1.8.2020: s.v. dependenssikielioppi. (Tarkka osoite: <https://www.tieteentermipankki.fi/wiki/Nimitys:dependenssikielioppi>.); Verkkokielioppi 1.8.2020 s.v. Selityksiä ja lisätietoja. (Tarkka osoite: <https://fl.finnlectura.fi/verkkosuomi/selityks.htm>)

⁸³ Koskeniemi 2002.

⁸⁴ Hajič et al. 2006.

⁸⁵ Bamman & Crane. 2006.

⁸⁶ Passarotti M. 2011.

Pragmatic Resources of Old Indo-European Languages⁸⁷ (PROIEL) -puupankin latinankielinen osuus koostuu lähinnä Vulgatasta, jonka lisäksi se sisältää muutamia antiikin ja myöhäisantiikin tekstejä (yhteensä 200 000 sanetta). Vaikka kaikki kolme skeemaa ovat saman tyyppisiä, niiden välillä on joitakin eroja sekä merkinnöissä että annotointiperiaateissa. Myös varhaiskeskiaikaisista asiakirjateksteistä (580 000 sanetta) koostuva Late Latin Charter Treebank⁸⁸ (LLCT) noudattaa samoja perusperiaatteita myöhäislaitin erityispiirteet huomioon ottavin lisäyksin. Näiden lisäksi Perseids Projectin (johon myös ALDT ja AGDT sisältyvät) puitteissa on otteita antiikin teksteistä sisältävä Tufts University Treebanked Commentaries⁸⁹ (ns. Harrington Trees), jonka annotointikehikko perustuu ALDT/AGDT:iin mutta käyttää yksityiskohtaisempia, perinteisten kielioppien luokitteluja hyödyntäviä merkintöjä.⁹⁰

Viimeistä lukuun ottamatta kaikista edellä mainituista puupankeista, tai ainakin niiden osista, on myöhemmin tehty myös UD-kehikon mukaiset versiot. Näissäkään annotointiratkaisut eivät ole yhteisistä yleisperiaateista huolimatta täysin yhdenmukaisia, ja epäjohdonmukaisuuksia esiintyy myös samojen puupankkien sisällä. Universal Dependencies -hankkeessa kehitetään useille kielille soveltuvaa johdonmukaista puupankkiannotointia. Annotointikehikossa syntaktisten riippuvuuksien merkitseminen perustuu Stanford dependencies -koodeihin⁹¹ ja sanaluokkaleimat Google universal part-of-speech -koodeihin⁹². Morfosyntaktisten kategorioiden pohjana on Interset interlingua⁹³. Puupankit esitetään Conference on Computational Natural Language Learning CoNLL-U -standardin mukaisessa muodossa. Projektin tavoitteena on yleispätevä valikoima kategorioita, ohjeita ja periaatteita, jotka soveltuvat annotointiin kielestä riippumatta ja mahdollistavat eri kielten samankaltaisten rakenteiden yhdenmukaisen merkitsemisen.⁹⁴

Vaikka sekä UD että PDT esittävät syntaktiset suhteet dependenssikieliopin mukaisesti, ne eroavat toisistaan muutamalla merkittävällä tavalla. Kehikoiden taustalla olevat kieliteoriat ja tavoitteet ovat osittain erilaista. PDT pohjautuu funktionaalis-generatiiviseen kuvaukseen, jossa annotoinnin syntaktinen taso rakentuu morfologiselle tasolle ja luo pohjan

⁸⁷ Haug & Jøhndal. 2008.

⁸⁸ Korkiakangas 2021.

⁸⁹ Harrington (käyty 31.7.2020) <https://perseids-publications.github.io/harrington-trees/>

⁹⁰ Cecchini et al. 2018.

⁹¹ esim. de Marneffe et al., 2014.

⁹² Petrov et al. 2012.

⁹³ Zeman, 2008.

⁹⁴ Nivre et al. 2016.

syvärakennetasolle (engl. *tectogrammatical layer*).⁹⁵ Pyrkimyksenä on mahdollisimman tarkka kielellinen kuvaus, minkä tuloksena annotointisäännöt voivat olla monimutkaisia ja kielikohtaiset erot suuria. UD perustuu leksikaalis-funktionaaliseen kielioppiin⁹⁶ ja pyrkii mahdollisimman yksinkertaisiin, kieliriippumattomiin annotointisääntöihin, joiden tuottamat puurakenteet soveltuvat mahdollisimman hyvin esimerkiksi hakutoimintoihin.⁹⁷

Käytännössä näkyvimvät erot UD- ja PDT-annotointikehikoiden välillä johtuvat siitä, että UD:n keskeisenä periaatteena on mahdollisimman matala puurakenne, jossa määreet liittyvät suoraan pääsanoihinsa ja lausekkeen pääsanana on aina sisältösana, kun taas PDT:ssa käytetään adpositioita, konjunktioita ja välimerkkejä siltarakenteina riippuvuussuhteiden kieliopillisesti mahdollisimman tarkan esityksen saavuttamiseksi. UD:n etuna on, että syntaktisesti analogiset rakenteet tulevat hyvin pitkälti merkityiksi johdonmukaisesti samalla tavoin. Vaarana kuitenkin on, että analyysin kannalta oleellista informaatiota katoaa. Tätä voidaan jossakin määrin helpottaa ”Enhanced Dependencies” -lisämerkinnöillä, joiden hyödynnettävyys nykyisillä työkaluilla (esim. hakutoiminnot) on tosin perusannotaatioita heikompaa. PDT:n etuna ovat kunkin kielen erityispiirteet huomioon ottavat käytännöt ja kieliopillinen tarkkuus, kuten ALDT:n latinankielisessä tekstissä tyypillisten elliptisten rakenteiden UD:n perusmallia täsmällisempi käsittely. Tarkkuuden varjopuolena on joidenkin rakenteiden ja merkintöjen jonkin asteinen mielivaltaisuus: esimerkiksi sivulauseen predikaatti liittyy päälauseeseen suoraan tai siltarakenteella (sivulauseen aloittavan konjunktin välityksellä) sivulauseen tyypistä riippuen, ja välimerkkejä käytetään tarpeen mukaan rinnasteisten ilmausten (koordinaatio ja appositio) apumerkintöinä.

Tässä työssä olen päättänyt käyttämään UD-kehikkoa monesta syystä. UD tuntuu olevan vakiinnuttamassa asemaansa tärkeimpänä dependenssikielioppiin perustuvana annotointijärjestelmänä, minkä ansiosta tätä tukevia valmiita, suhteellisen helppokäyttöisiä tietoteknisiä työkaluja on saatavilla kohtuullinen valikoima. Kuten edellä luvussa 4.3 on esitetty, manuaalisesti annotoitua puupankkia voidaan automaattisten jäsentimien ja vapaasti saatavilla olevien jäsennymallien ansiosta täydentää laajemmalla, tietoteknisin välinein tuotetulla aineistolla (jonka laatu tosin vaihtelee). UD:n tietyt annotointiratkaisut – kuten kopulan käsittely ja johtolauseiden merkitseminen dialogimuotoisessa tekstissä – soveltuvat ALDT:tä paremmin käsiteltävään tekstiin. Myös ajatus puupankin soveltuvuudesta kieltenväliseen vertailuun on mahdollisen jatkokäytön näkökulmasta kiinnostava, sillä

⁹⁵ esim. Sgall et al. 1986.

⁹⁶ Bresnan 2001.

⁹⁷ Rosa et al. 2014; Nivre et al. 2016.

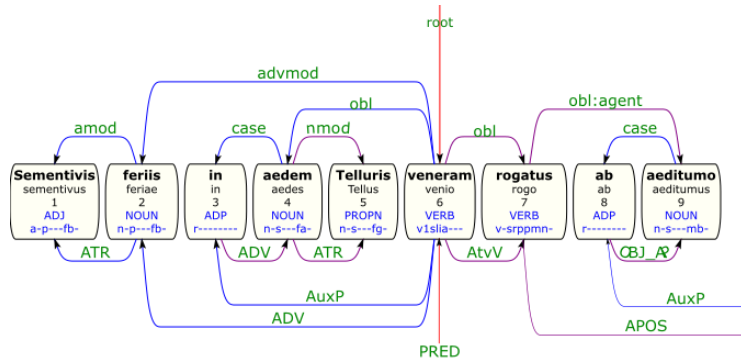
käsiteltävä teksti on viimeistelemätöntä ja saattaa sisältää esimerkiksi puhekielelle ominaisia piirteitä, joiden vertailu myöhempiin (romaanisiin) kielimuotoihin voisi olla mielenkiintoista.

Seuraavaksi havainnollistan käsiteltävästä aineistosta poimituin esimerkkilausein em. annotointikehikkojen välisiä eroja sekä syntaktisen annotoinnin keskeisiä ongelmakohtia ja niihin tässä työssä valittuja ratkaisuja. Esimerkit ovat käsiteltävän tekstin dialogiosan alusta⁹⁸, josta olen annotoinut noin 50 lauseen pituisen osan sekä UD- että ALDT-kehikon mukaisesti. Puukuvioissa yläpuolella olevat merkinnät ovat UD:n mukaisia, ja alapuolella vastaavasti ALDT:n.

Suorat pääsana-määritesuhteet vs. siltarakenteet

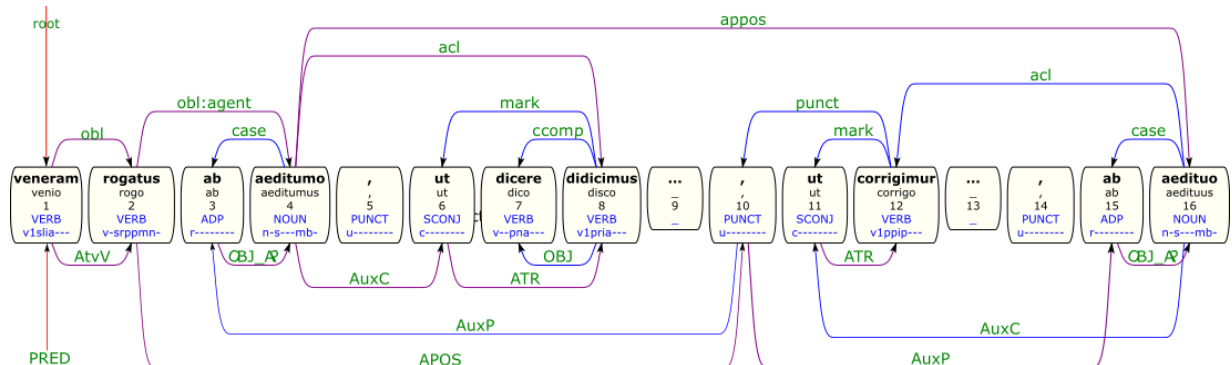
Kuvion 3 lauseessa predikaattia *veneram* määrittää paikan adverbiaalilauseke *in aedem Telluris*. UD:ssä määritelausekkeen pääsana *aedem* liittyy suoraan omaan pääsanaansa eli predikaattiin ja prepositio *in* liittyy vastaavasti lausekkeen pääsanaan *case*-relaatiolla. ALDT:ssä predikaattiin liittyy prepositio *in*, johon lausekkeen sisällöllinen pääsana liittyy syntaktista funktiotaan vastaavalla ADV-relaatiolla.

Kuvio 3: *Sementivis feriis in aedem Telluris veneram rogatus ab aeditumo (...)* (Rust 1,2,1, alku)



Vastaavalla tavalla kuviossa 4 sivulauseeseen *ut dicere didicimus* pääsana liittyy omaan pääsanaansa *aeditumo* UD:ssä suoraan *acl*-relaatiolla, ALDT:ssä *ut*-konjunktion välityksellä.

Kuvio 4: *(...) rogatus ab aeditumo, ut dicere didicimus (...), ut corrigimur (...), ab aedituo* (Rust 1,2,1, loppu)

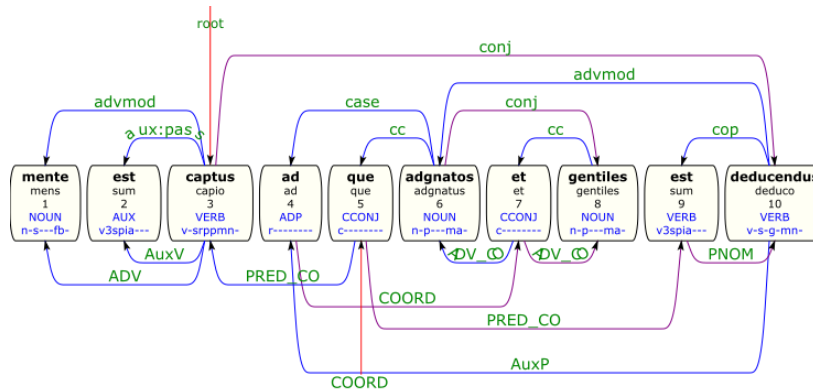


⁹⁸ Rust 1.2.1.

Rinnasteiset ilmaukset

Annotointikäytäntöjen erot tulevat erityisen merkittävällä tavalla esiin useamman syntaktisesti samassa tehtävässä olevan ilmauksen (koordinaatio ja appositio) merkitsemisessä. Edellä olevassa esimerkissä (kuvio 4) passiivin partisiippiä *rogatus*⁹⁹ määrittää kaksi keskenään appositiossa olevaa agenttirakennetta *ab aeditumo* ja *ab aedituo*. UD:ssa varsinainen syntaktinen relaatio *obl:agent* merkitään pääsanan ja ensimmäisen appositiolausekkeen välille, ja muut appositiolausekkeet puolestaan liittyvät ensimmäiseen *appos*-relaatiolla. Vastaavasti kuvion 3 kahdesta rinnakkaisesta predikaatista *(est) captus* ja *(est) deducendus* ensimmäinen on merkitty *root*-elementiksi, johon jälkimmäinen liittyy *conj*-relaatiolla. Samassa lauseessa jälkimmäiseen verbiin liittyy vastaavalla tavalla myös kaksi adverbiaalilauseketta *ad adgnatos* ja *(ad) gentiles*. Kaikki rinnastuskonjunktiot (tässä *-que* ja *et*) liittyvät pääsanoihin *cc*-relaatiolla ja kaikki välimerkit *punct*-relaatiolla.

Kuvio 5: *mente est captus adque adgnatos et gentiles est deducendus* (Rust. 1,2,8)



ALDT:ssä (ja yleensä PDT:ssä) ilmausten välissä olevaa välimerkkiä (kuten kuvion 4 lauseessa pilkkua) tai rinnastuskonjunktia käytetään erityisenä koordinaatiosolmuna, joka liittyy pääsanaan *APOS*- (kuvio 4) tai *COORD*-relaatiolla (kuvio 5), ja johon puolestaan liittyvät kaikki appositiossa olevat tai rinnasteiset osat. Apposition tai koordinaation osat merkitään suhdetta pääsanaan ilmaisevaan relaatioon liitettävällä *_AP*- tai *_CO*-lisäleimalla (kuviossa 2 *OBJ_AP*, kuviossa 3 *PRED_CO* ja *ADV_CO*).

Kuvion 5 rinnasteiset adverbiaalimääreet havainnollistavat UD:n käytännöstä seuraavaa informaation hukkaamista: merkinnöistä ei voida päätellä, liittyykö prepositio *ad* vain koordinaation ensimmäiseen osaan *adgnatos* vai myös jälkimmäiseen *gentiles*. ALDT:ssä tämä on yksiselitteistä, sillä prepositio liittyy koordinaatiosolmuun *et* ja siten kaikkiin koordinaatiossa mukana oleviin ilmauksiin. Erityistapauksen muodostaa puun juuri, joka

⁹⁹ Perinteisissä kieliopeissa myös tällaista rakennetta (*rogatus*, ”pyytämänä”) voidaan kutsua appositioiksi – tällaisten ilmausten määritelmiä ja annotointikäytäntöjä käsitellään vielä jäljempänä tarkemmin.

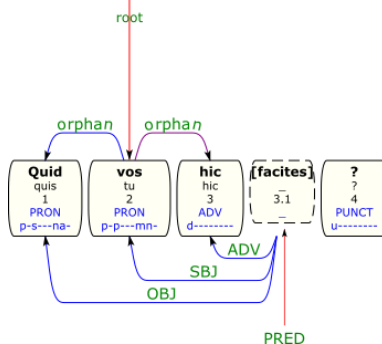
UD:ssa on **relaatio** *root*. ALDT:ssä puu alkaa *ROOT*-nimisestä **solmusta**, joka ei esitysteknisistä syistä näy näissä esimerkkikuvioissa. Ensimmäinen varsinainen elementti liittyy *ROOT*-solmuun yleensä *PRED*-relaatiolla mutta koordinaatiotapauksessa *COORD*-relaatiolla, kuten kuviossa 3.

Ellipsi

Myös elliptisten ilmausten käsittely erottaa annotointikehikot selvästi toisistaan ja on omiaan aiheuttamaan tulkinnallisia epäselvyyksiä. Koska UD:n perusperiaatteena on mahdollisimman matala puurakenne ja määreiden liittyminen pääsanaansa mahdollisimman suoraan, pois jätettyjä elementtejä ei yleensä merkitä millään tavalla erikseen vaan tämän paikan ottaa puurakenteessa seuraava jäsen sovitussa järjestyksessä. Esimerkiksi pois jätetyn predikaatin paikan ottaa ensisijaisesti siihen liittyvä subjekti, kuten *vos* kuvion 6 yksinkertaisessa kysymyslauseessa.

Käytännön ilmeisenä varjopuolena on, että lauseen muut osat eivät ole mielekkäässä syntaktisessa suhteessa pois jätetyn elementin korvaajaan, varsinkin niissä sangen yleisissä tapauksissa, joissa pois jätetty osa on lauseen predikaattiverbi. Nämä liitetäänkin korvaajasaneeseen erityisellä *orphan*-relaatiolla. ALDT:ssä elliptiset elementit lisätään puuhun ns. tyhjinä solmuina (empty node), kuten esimerkklauseessa predikaatti *facitis* (tms.), jolloin lauseen syntaktiset suhteet voidaan esittää täsmällisesti. Vastaava analyysi on mahdollista esittää UD:ssa käyttäen *enhanced dependencies* -relaatioita.

Kuvio 6: *Quid vos hic [facites]? (Rust 1,2,1)*

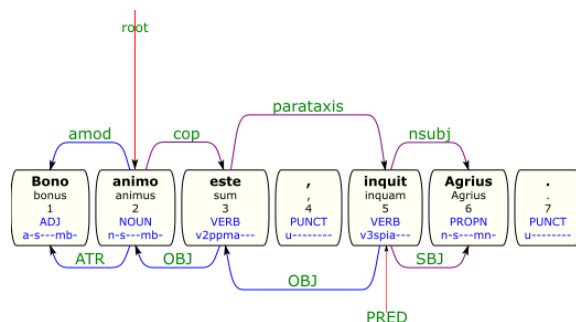


Yksi tyypillinen tapaus predikaatin ellipsistä ovat rinnasteiset lauseet, joiden yhteinen predikaatti on usein ilmaistu vain ensimmäisessä lauseessa. Yhdessä edellä esitetyn koordinoitujen ilmausten merkitsemiskäytäntöjen kanssa nämä saattavat tuottaa hyvin hankalasti hahmotettavia puurakenteita. Toinen yleisesti pois jätetty elementti on kopula tai apuverbi, joita käsitellään seuraavaksi tarkemmin.

Kopula ja apuverbi

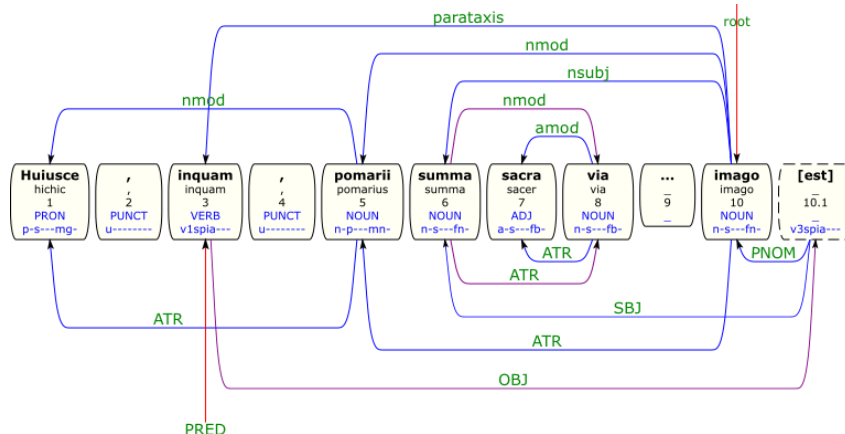
Kummassakin annotointikehikossa apuverbi analysoidaan pääverbin määreeksi, kuten kuvion 3 lauseen passiivirakenne *est captus* (UD: *aux:pass*, ALDT: *AuxV*). Predikatiivirakenteissa ALDT sen sijaan pitää kopulaa verbilausekkeen pääsanana johon predikatiivi liittyy PNOM-relaatiolla, kuten samassa lauseessa *est deducendus*. UD kohtelee kopulaa analogisesti apuverbin kanssa ja analysoi tässä tapauksessa pääsanaksi gerundiivin, jonka määre kopula (relatio *cop*) on. Samoin kohdellaan mitä tahansa predikatiiveja, eli predikaattinominia (tai koko ”jonakin olemisen” lauseketta) käsitellään kuten verbiä. Esimerkiksi kuviossa 7 lausekkeen *animo este* pääsanaksi (ja tässä koko virkkeen juureksi) merkitään *animo*.

Kuvio 7: *Bono animo este, inquit Agrius.* (Rust 1,2,11)



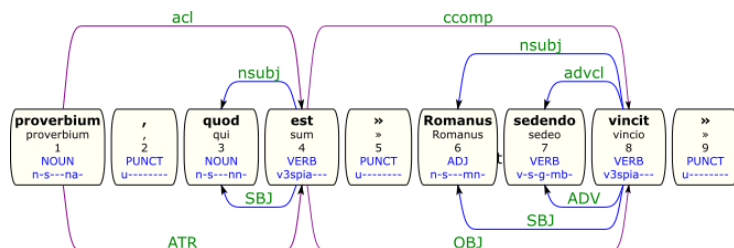
Tämä UD:n *esse*-verbin syntaksin kannalta analoginen analysointikäytäntö on nähdäkseni johdonmukaisempi kuin ALDT:n tapa, jossa relaation suunta ja jopa kohde riippuu osittain semanttisesta tulkinnasta. Esimerkiksi pohdinta siitä, pitäisikö sellaisissa nesessiivirakenteissa kuin edellä mainittu *est deducendus* (ns. passiivin perifrastinen muoto), *esse* tulkita kopulan sijaan pikemminkin apuverbiksi (näin ei käytännössä tehdä), ei vaikuta puun rakenteeseen. Toinen etu annotoinnin selkeyden ja sujuvuuden kannalta tässä käytännössä tulee siitä, että juuri kopulan pois jääminen on latinankielisissä teksteissä erittäin yleistä. UD:n merkintätavalla tästä ei kuitenkaan synny varsinaista ellipsiä, joka aiheuttaisi informaatiokatoa tai muuta tulkinnallista epäselvyyttä. Esimerkkinä viimeksi mainitusta on *imago [est]* kuviossa 8.

Kuvio 8: *huiusce, inquam, pomarii summa sacra via (...) imago [est]* (Rust 1,2,10)



Täysin aukottoman johdonmukaiseen esitystapaan ei UD kuitenkaan tässä pääse, sillä (verraten harvinaisissa) tapauksissa, joissa predikatiivin asemassa on kokonainen lause, yksiselitteinen merkintätapa tällä periaatteella ei ole mahdollinen. Tällöin joudutaan tyytymään ratkaisuun, joissa kopula ottaa pääsanana paikan, kuten kuvion 9 suorassa lainauksessa.

Kuvio 9: *proverbium, quod est "Romanus sedendo vincit"* (Rust 1,2,2)



Johtolauseet (parataksi)

Yllä olevien kuvioiden 7 ja 8 esimerkeistä näkyy, että johtolauseiden analysointi eroaa UD- ja ALDT-käytännöissä toisistaan. UD:ssa johtolauseet (*inquit Agrius; inquam*) analysoidaan löyhästi muuhun lauserakenteeseen liittyviksi rinnasteisiksi, parataktisiksi suhteiksi, jotka merkitään *parataxis*-relaatiolla koko virkkeen pääsanana määreiksi. ALDT:ssä käytäntö pohjaa tässäkin osittain tekstin merkitykseen, ja suora esitys analysoidaan samoin kuin vastaava epäsuora, jolloin johtolauseen predikaatti on koko virkkeen juuri. Erityisesti tässä työssä tutkittavana olevan dialogimuotoisen tekstin esittämiseen UD:n tapa on huomattavasti käyttökelpoisempi¹⁰⁰, koska kertojan ja puhujien tekstit erottuvat puupankissa selkeästi, eikä repliikin analysointi riipu siitä, sisältääkö se tekstissä johtolauseen.

¹⁰⁰ Olisi tietysti mahdollista myös soveltaa ALDT:tä ohjeista tältä osin poiketen, kuten on tehty mm. Phaedrussuupankin (<http://www.perseids.org/tools/arethusa/app/#/perseids?chunk=1&doc=27645>) kohdalla.

Appositio ja nominien määreet

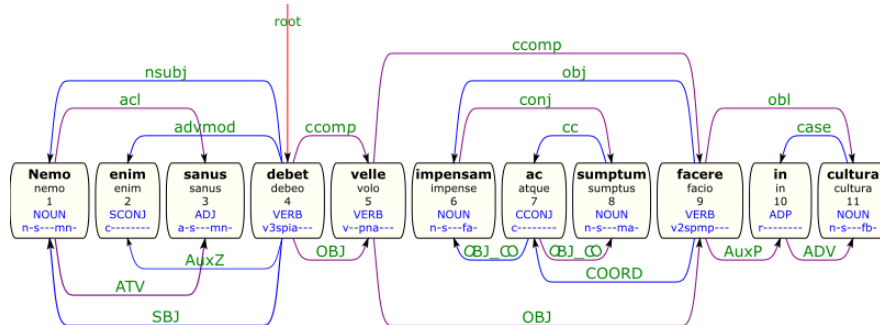
Edellä rinnastuksen yhteydessä on käsitelty appositionrakenteiden annotoimista. Kummassakin annotointikehikossa appositio viittaa vain samassa lauseenjäsenetävässä oleviin lausekeisiin, joista mikään ei määritä toisia. Perinteisissä (latinan) kieliopissa appositioiksi kutsutaan myös muita määreitä, jotka tyypillisesti kongruoivat subjektin kanssa mutta samalla määrittävät myös verbiä. Usein nämä vastaavat merkitykseltään suomen ns. predikatiiviadverbiaalia. Koska molemmissa annotointikehikoissa edellytetään, että määreellä on vain yksi pääsana, näiden rakenteiden analysoinnissa on tehtävä valintoja, joiden seurauksena informaatiota katoaa.

Kun tällaisena ”appositiona” on partisiippi (ns. *participium coniunctum*), dependenssisuhteet on melko suoraviivaista määrittää yleisten annotointisääntöjen pohjalta. Substantiivi- ja adjektiivimääreet tyyppiä *Cicero consul* (merkityksessä *konsulina Cicero*) tai *Cato senex* (*vanhana Cato*) on sen sijaan ratkaistu eri annotointikehikoissa eri tavoin, ja myös samoja ohjeita noudattavien puupankkien välillä (joskus jopa sisällä) on merkintäkäytännöissä vaihtelua. ALDT:ssa on tällaisia relaatioita varten erilliset merkinnät: Jos pääsana (subjekti) on näkyvissä, kuten *Nemo* kuviossa 10, määre (tässä *sanus*) liittyy siihen ATV-relaatiolla. Jos ilmisubjekti puuttuu, kuten on latinassa hyvin usein, määre liitetään verbiin merkinnällä *AtvV*, esimerkiksi *veneram rogatus* kuviossa 3. UD:ssa tällaiselle relaatiolle ei ole omaa merkintää, vaan ne käsitellään – analogisesti yllä mainittujen vastaavien partisiippirakenteiden kanssa – adjektiivilauseina ja liitetään substantiivipääsanaansa merkinnällä *acl*¹⁰¹. Ilmisubjektin puuttuessa määre liittyy verbiin tavallisesti *obl*-relaatiolla, mutta käytäntö latinankielisissä UD-puupankeissa ei ole yhtenäinen. Perseus- ja ITTB-puupankkien UD-versioissa *AtvV*-relaatiot näyttävät olevan säännönmukaisesti korvattu *xcomp*-relaatiolla, jonka käyttö ohjeiden¹⁰² mukaan on kuitenkin yksiselitteisesti rajattu tapauksiin, jossa määre on syntaktisesti pakollinen.

¹⁰¹¹⁰¹ puhdas adjektiiviattribuutti sen sijaan *amod*.

¹⁰² <https://universaldependencies.org/u/dep/xcomp.html>

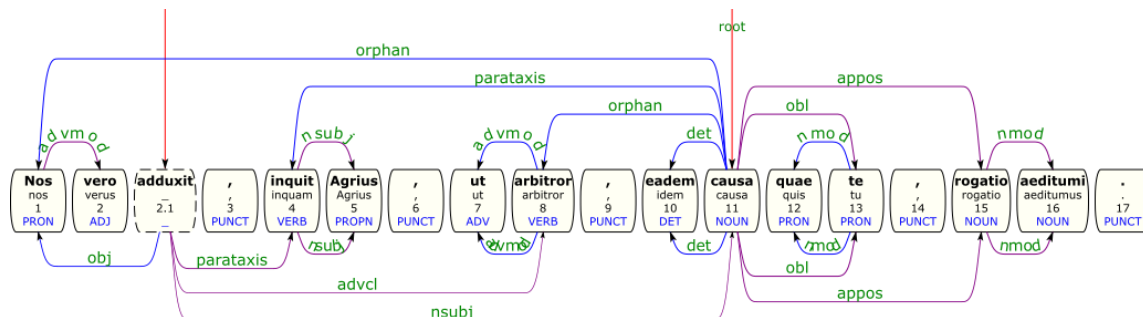
Kuvio 10: *Nemo enim sanus debet velle impensam ac sumptum facere in cultura* (Rust 1,2,8)



Enhanced Dependencies

Kuvio 11 havainnollistaa, että laajennettujen riippuvuussuhteiden merkitsemiseen tarkoitettuja *enhanced dependencies* -relaatioita (kuvion alaosa) käyttämällä voidaan useita UD-kehikon puutteita ja ongelmia paikata. Tässä esimerkissä on elliptinen predikaatti *adduxit* (2.1) lisätty lauseeseen, jolloin perusversion juuri *causa* voidaan jäsentää subjektiksi ja tähän *orphan*-relaatiolla liittyvät määreet saavat syntaktista asemaansa kuvaavat relaatiot. Tämän työn puitteissa *enhanced dependencies* -relaatioiden laajempi käyttö ei kuitenkaan ole ollut mahdollista.

Kuvio 11: *Nos vero [adduxit], inquit Agrius, ut arbitror, eadem causa quae te, rogatio aeditumi* (Rust 1,2,2)



Ut-rakenteiden analysointi ja annotointi

Tässä työssä tarkastellaan erityisesti *ut*-rakenteiden käyttöä dialogin eri puhujilla. Tällaisia rakenteita on sekä syntaksiltaan että merkitykseltään monenlaisia ja ne ovat olleet laajasti analyysin ja tutkimuksen kohteina jo antiikin kieliopeista alkaen. Perinteisesti *ut*-rakenteet voidaan jakaa pääluokkiin kahdella perusteella: Ensinnäkin monet sanakirjat (ja edellä esitetyt annotointikehikot) erottavat toisistaan *ut*-adverbin ja *ut*-konjunktion. Toiseksi ilmaisut, joissa *ut* on luokiteltava sanaluokaltaan konjunktioksi, voidaan jakaa substantiivi- ja adverbialilauseisiin (tai -lausekkeisiin) tai pakollisiin (komplementti, argumentti) ja vapaisiin (adjunkti) määreisiin.¹⁰³

¹⁰³ Yksityiskohtainen keskustelu aiheesta Bolkestein 1976.

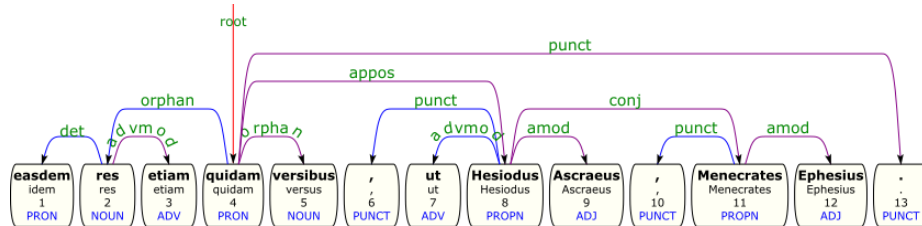
Adverbinä *ut* voi olla joko tapaa ilmaiseva (merkityksessä ”siten”, ”kuten”), niin kuin edellä kuviossa 11 *ut arbitror*, tai interrogatiivinen (”kuinka”). Adverbin ja (alistus)konjunktion raja ei tässä ole aina ilmeinen ja niin kielioppitraditiot kuin annotointikäytännötkin poikkeavat toisistaan.¹⁰⁴ Tässä työssä olen päätenyt käytäntöön, jossa *ut* jäsennetään adverbiksi vain sellaisissa tapauksissa, joissa se ei ole syntaktisesti välttämätön, eli ainoastaan tuo lisätietoa verbin merkityksestä. Toisin sanoen näissä tapauksissa *ut*-sanana voisi jättää pois ilman, että rakenteesta tulee epäkieliopillinen. Näin ollen esimerkiksi kuvion 4 lauseissa *ut dicere didicimus* ja *ut corrigimur*, on *ut* sivulauseen aloittava konjunktio, vaikka semanttista eroa edellä mainittuun *ut arbitror* -tapaukseen ei juuri ole.

Kieliopit listaavat koko joukon erityyppisiä, *ut*-konjunktioilla alkavia sivulauseita. Luokitteluperusteena on tällöin yleensä merkitys: seurausta ilmaisevat konsekutiivilauseet, tarkoitusta ilmaisevat finaalilauseet, vertailevat komparatiivilauseet, myönnytystä ilmaisevat konsessiivilauseet ja ajan ilmauksiin liittyvät temporaalilauseet. Syntaktisen jäsennyksen ja annotoinnin näkökulmasta merkitystä keskeisempää on kuitenkin määrään pakollisuus, joka ei suoraan noudata merkitykseen perustuvaa jakoa. Esimerkiksi konsekutiiviset *ut*-lauseet voivat olla joko subjektin (UD-relaatio *csubj*) tai objektin (*ccomp*) asemassa olevia substantiivilauseita tai vapaaehtoisia adverbilauseita (*advcl*).

Olemassa olevissa puupankeissa on häilyvyyttä lähinnä komparatiivisiksi luettavien *ut*-ilmausten annotoinnissa. Kuvion 12 lauseessa *ut* määrittää apposiitioksi tulkittavaa pääsanaa adverbina (merkityksessä *kuten*, *esimerkiksi*, *nimittäin* tms.). Perseus- ja ITTB-puupankkien UD-versioissa valtaosa *ut*-adverbin määrittämisestä lauseista on jäsennetty pakollisiksi *xcomp*-relaatiolla, mikä on nähdäkseni useissa tapauksissa ohjeiden vastaista, kun kyse on selkeästi vapaasta adverbialilauseesta. Toisaalta itse *ut*-sanana jäsennys adverbiksi tai konjunktioiksi on häilyvää, ja Perseus-puupankissa *ut* on joissakin tapauksissa jopa liitetty pääsanaansa *case*-relaatiolla, mikä todennäköisesti ei kuitenkaan ole tarkoituksellista.

¹⁰⁴ Adverbin määritelmä ei ole yksiselitteinen. Tällainen *ut comparativum* ei ole uudempien kielioppien mukainen varsinainen adverbi. Tässä esitetty luokittelu vastaa annotointikehikoiden dokumentaatiossa viitattuihin, lähinnä englanninkielisten latinan kieliopin ja sanaston esityksiin. Suomalaisissa esityksissä Linkomies (1933:98) puhuu vertailupartikkelista, Pekkanen (1981:163) komparatiivisista sanoista. Toisaalta ranskalainen Gaffiot-sanakirja (2016) listaa komparatiivisen merkityksen vain adverbina.

Kuvio 12: *eadem res etiam quidam versibus, ut Hesiodus Ascraeus, Menecrates Ephesius.* (Rust 1,1,9)



6 TULOKSET

Tässä luvussa esittelen aineiston kvantitatiivisen analyysin tuloksia. Ensin tarkastelen taustaksi tekstin dialogirakennetta ja puhujien rooleja puheenvuorojen määrän ja pituuden valossa. Puhujien välisiä eroja funktiosanojen frekvensseissä ja lauserakenteissa analysoidaan automattisesti jäsennetyn tekstin perusteella, ja tässä mielessä osa tuloksista on alustavia ja menetelmien käyttöä havainnollistavia.

6.1 Tekstin rakenne dialogina

Kuten edellä luvussa 3.1 on mainittu, *Res Rusticae* dialogeissa esiintyy vaihteleva määrä pää- ja sivuhenkilöitä erilaisissa rooleissa. Kolmen kirjan dialogit eroavat toisistaan paitsi henkilöiden määrän, myös puheenvuorojen vaihtelun ja ”tempon” osalta. Taulukossa 1 on esitetty puhujien saneiden ja puheenvuorojen määrät eri kirjoissa. Sanemäärät on tässä laskettu ilman välimerkkejä. Ensimmäisessä kirjassa 14 480 sanetta jakautuu 65 puheenvuoroon, joiden keskipituus on näin ollen 233 sanetta. Toisessa kirjassa on 10 927 sanetta ja 40 keskimäärin 273 saneen puheenvuoroa. Kolmannessa kirjassa saneita on 10 228 ja puheenvuoroja 52, eli keskimäärin 197 sanetta puheenvuoroa kohti. Jo näiden lukujen perusteella voidaan nähdä, että toisessa kirjassa puheenvuorot ovat pidempiä ja niitä on vähemmän kuin kahdessa muussa kirjassa.

Taulukko 1: Puheenvuorojen ja saneiden määrät eri puhujilla

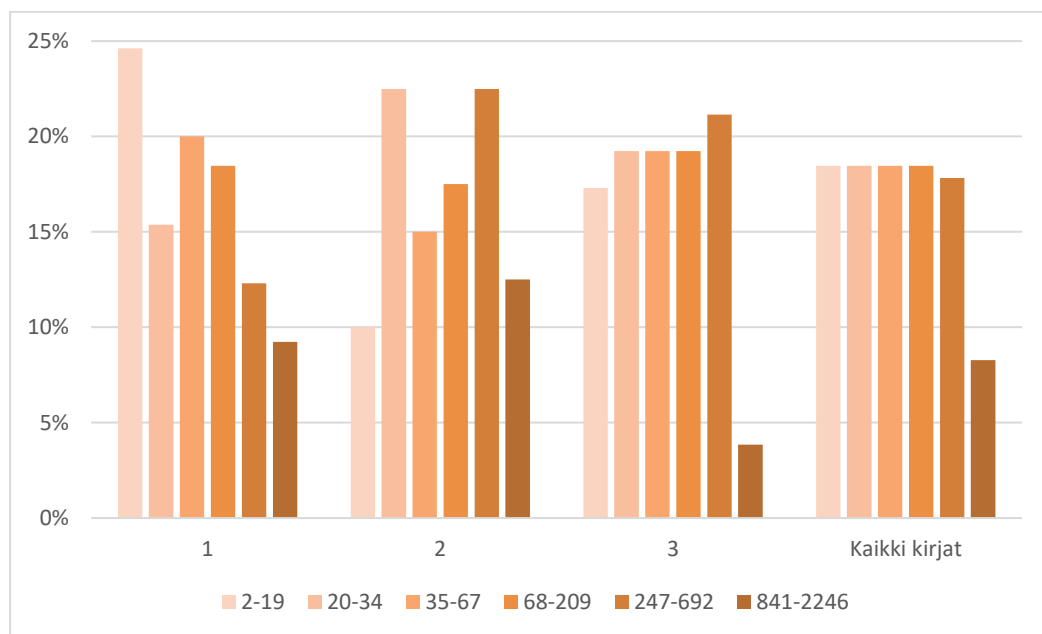
1. kirja				2.kirja				3.kirja			
Henkilö	Puheenvuoroja	Saneita	Saneita / puheenvuoro	Henkilö	Puheenvuoroja	Saneita	Saneita / puheenvuoro	Henkilö	Puheenvuoroja	Saneita	Saneita / puheenvuoro
Varro/kertoja	9	1153	128	Varro/kertoja	15	1609	107	Varro/kertoja	11	1369	124
Scrofa	15	7061	471	Scrofa	4	2480	620				
Stolo	15	4872	325	Atticus	5	2308	462	Merula	13	5755	443
Fundanius	7	800	114	Cossinius	7	1867	267	Appius	8	1754	219
Agrius	14	479	34	Lucienus	2	1004	502	Axius	18	1266	70
Agrasius	5	115	23	Vaccius	3	943	314	Pica	1	56	56
				Murrius	2	657	329	Pavo	1	28	28
Yhteensä	65	14480	223	Yhteensä	40	10927	273	Yhteensä	52	10228	197
Kakki kirjat yhteensä											
Varro/kertoja	35	4131	118								
Scrofa	19	9541	502								
Kaikki yht.	157	35635	227								

Puheenvuorojen määrä ja keskipituus antavat kuitenkin vain hyvin karkean kuvan eroista, joita voidaan havaita puheenvuorojen jakautumisessa dialogien välillä. Olen jakanut kaikki

kolmen kirjan yhteensä 157 puheenvuoroa niiden pituuden perusteella kuuteen ryhmään. Yksi ryhmä käsittää 13 kaikkein pisintä, yli 840 saneen puheenvuoroa. Loput viisi ryhmää ovat tasakokoisia niin, että kussakin on 28 tai 29 puheenvuoroa.

Kuviossa 13 on esitetty näiden puheenvuororyhmien jakauma kirjoittain. Voidaan havaita, että kolmannessa kirjassa eri pituiset puheenvuorot jakautuvat hyvin samalla tavoin kuin koko aineistossa keskimäärin lukuun ottamatta kaikkein pisimpiä puheenvuoroja, joita on vain pari kappaletta. Ensimmäisen kirjan puheenvuoroissa on muihin kirjoihin verrattuna paljon hyvin lyhyitä repliikkejä ja vastaavasti vähän pitkiä puheenvuoroja. Toisaalta ensimmäisissä kirjassa on muutama erittäin pitkä puheenvuoro (1427, 1591, ja 2246 sanetta), jotka yhdessä kattavat yli kolmasosan kirjan koko pituudesta. Toisessa ja kolmannessa kirjassa on kummassakin vain yksi noin 1500 saneen puheenvuoro. Toisessa kirjassa leimallista on välihuomautuksen omaisten lyhyiden repliikkien vähäisyys ja kahta muuta kirjaa suurempi osuus pitkähköjä puheenvuoroja (ryhmä 247–692 sanetta sekä ryhmästä 841–2246 sanetta alle 1000 saneen puheenvuoroja).

Kuvio 13: Eri pituisten (sanemäärä) puheenvuorojen prosenttiosuudet kirjoittain



Puheenvuorojen vaihtelusta ja keskustelun temposta saa yksityiskohtaisemman kuvan liitteen 4 kaavioista, joissa puheenvuorot on esitetty sanemäärään perustuvalla aikajanalla. Kaikki kolme kirjaa alkavat Varron esipuheella, joka ei kuulu itse dialogiin. Ensimmäisessä ja kolmannessa kirjassa esipuhetta seuraa useista lyhyistä puheenvuorosta koostuva, johdatteleva keskusteluosuus. Tämän jälkeen suurin osa puheenvuoroista on pitkiä, luentomaisia, mutta näiden välissä on lyhyehköjä muutaman repliikin siirtymäjaksoja.

Toisessa kirjassa sitä vastoin johdantokeskustelu on lyhyt, ja dialogia leimaavat keskipitkät puheenvuorot ja useamman puhujan osallistumien keskusteluun vuorollaan. Tässäkin kirjassa on tosin muutamia lyhyiden repliikkien siirtymäjaksoja.

Puhujittain tarkasteltuna taulukosta 1 huomataan, että ensimmäisessä ja kolmannessa kirjassa on kaksi ”luennoitsijaa”, joiden puheenvuorojen keskimääräinen sanemäärä on muita dialogin henkilöitä selvästi suurempi. Ensimmäisessä kirjassa tällaisia henkilöitä ovat Scrofa ja Stolo, kolmannessa Merula ja Appius. Toisessa kirjassa sen sijaan dialogin kaikkien henkilöiden puheenvuorot ovat keskimäärin melko pitkiä. Varron omien puheenvuorojen keskimääräistä sanemäärää kaikissa kirjoissa nostavat esipuheet, jotka ovat 400–700 saneen pituisia. Toisaalta etenkin toisessa kirjassa Varron kertojan roolissa esittämät välihuomautukset lyhentävät hänen puheenvuorojensa keskimääräistä pituutta.

Taulukossa 2 pituuden mukaan ryhmitellyt puheenvuorot on esitetty puhujittain. Edellä havaitut erot kirjojen välillä näkyvät vielä selvemmin: Ensimmäisessä ja kolmannessa kirjassa kaksi asiantuntijaa on selvästi erikoistunut pitkiin puheenvuoroihin, ja sivuhenkilöillä on suuri määrä lyhyitä puheenvuoroja. Toisessa kirjassa kaikilla puhujilla on tasaisemmin eri pituisia puheenvuoroja. Ainakin tässä mielessä ensimmäinen ja kolmas kirja muistuttavat rakenteeltaan esimerkiksi Ciceron filosofisia dialogeja hyvin pienessä roolissa olevine sivuhenkilöineen, mutta toisessa kirjassa varsinaisia sivuhenkilöitä ei ole.

Taulukko 2: Dialogien henkilöiden puheenvuorojen jakautuminen pituuden (sanemäärä) mukaan

	Puheenvuoron sanemäärä						Yhteensä
	2-19	20-34	35-67	68-209	247-692	841-2246	
1. kirja							
Varro/kertoja	1	2	3	2	1		9
Scrofa	1	1	2	2	6	3	15
Stolo	3	1	4	4		3	15
Fundanius	2	2		2	1		7
Agrasius, Agrius	9	4	4	2			19
<i>Yhteensä, 1. kirja</i>	<i>16</i>	<i>10</i>	<i>13</i>	<i>12</i>	<i>8</i>	<i>6</i>	<i>65</i>
2. kirja							
Varro/kertoja	3	3	5	2	2		15
Scrofa		1	1			2	4
Atticus		1		1	2	1	5
Cossinius		1		3	3		7
Lucienus				1		1	2
Vaccius	1	1				1	3
Murrius					2		2
Epäselvä		2					2
<i>Yhteensä, 2. kirja</i>	<i>4</i>	<i>9</i>	<i>6</i>	<i>7</i>	<i>9</i>	<i>5</i>	<i>40</i>
3. kirja							
Varro/kertoja	3	2	4		2		11
Merula	2		1	3	5	2	13
Appius	1			4	3		8
Axius	3	7	4	3	1		18
Pavo, Pica		1	1				2
<i>Yhteensä, 3. kirja</i>	<i>9</i>	<i>10</i>	<i>10</i>	<i>10</i>	<i>11</i>	<i>2</i>	<i>52</i>
Yhteensä, kaikki kirjat	29	29	29	29	28	13	157

6.2 Funktiosanojen frekvenssejä

Seuraavassa tarkastelen aineiston yleisimpien lemموjen esiintymistä eri puhujilla. Rajaan tarkastelun tässä aineiston sisäiseen, puhujien väliseen vertailuun, vaikka Varron kielen yleisemmän arvioinnin kannalta vertailu myös klassisen latinan laajempaan korpukseen olisi mielenkiintoista. Aineistona on automaattisesti UDPipe-työkalulla Perseus-mallia käyttäen jäsennetty teksti. Käytettävissä olevat mallit tuottavat hieman erilaisia tuloksia, eikä parhaan tuloksen määrittäminen ole yksiselitteistä, ennen kuin riittävän suuri osuus aineistosta on annotoitu myös manuaalisesti. Perseus-malli soveltuu nimenomaan lemmaukseen kuitenkin todennäköisesti parhaiten, koska se sisältää pelkästään klassisen latinan tekstejä, kun taas muissa malleissa huomattava osa korpusta koostuu myöhäislatinan ja keskiajan aineistoista. Koko aineistossa vähintään sata kertaa esiintyvät lemmat ja niiden esiintymät kirjoittain on listattu liitteessä 5.

Tässä ensisijaisesti tarkasteltavien funktiosanojen esiintymäfrekvenssien luotettavuutta parantaa se, että monet automaattiseen lemmaukseen liittyvät ongelmat ovat näiden sanojen kohdalla pienempiä kuin tyypillisesti sisältösanoja tarkasteltaessa. Pronomineja ja apuverbiä lukuun ottamatta funktiosanat ovat taipumattomia. Konjunktoiden kohdalla homonymia ja polysemia ovat mahdollisia ongelmia lähinnä *quod*- ja *cum*-sanojen kohdalla: Edellinen voi

olla joko alistuskonjunktio tai *qui*-pronominin neutrimuoto, jälkimmäinen taas joko alistuskonjunktio tai prepositio. Liitteen 5 listassa *cum*-sanana polysemiaa ei ole otettu huomioon, vaan ilmoitetut esiintymät sisältävät sekä konjunktioita että prepositioidia. Sen sijaan seuraavissa taulukoissa 3 ja 5 nämä esiintymät on jaettu automaattisen jäsennyksen lauseenjäsenleiman perusteella.

Kuten edellä luvussa 5.2 on esitetty, *ut*-sana voi olla joko alistuskonjunktio tai adverbii. Tämä ero on kuitenkin tulkinnanvarainen eikä sitä ole tässä tarkastelussa tarpeen tehdä. Lisäksi konjunktioita vaihtoehtoinen muoto *uti* voi sekoittua samannäköiseen verbin infinitiiviin, mutta tästä aiheutuvat jäsennysvirheet eivät liene tulosten tulkinnan kannalta kovin vakavia. Pronominien *qui*, *is* ja *hic* taivutusmuotojen sekaantuminen samannäköisiin konjunktioihin, adverbeihin ja verbimuotoihin on todennäköisesti yleisemmin automaattisessa jäsennyksessä tapahtuva virhe, joka voi vääristää tässä esitettäviä tuloksia. Näidenkin kohdalla voitaneen kuitenkin melko turvallisesti olettaa, että jäsennysvirheet ovat satunnaisia eivätkä siis systemaattisesti vääristä vertailua eri puhujien välillä.

Alistuskonjunktioita

Taulukossa 3 on viiden yleisimmän aineistossa esiintyvän alistuskonjunktioita – *ut*, *quod*, *ne*, *cum* ja *si* – lukumäärät ja suhteelliset osuudet eri kirjoissa ja puhujilla. Alistuskonjunktioita frekvenssien tarkastelu on kielellisten piirteiden kannalta erityisen mielenkiintoista, sillä niiden käyttö saattaa antaa viitteitä myös puhujan käyttämistä lauserakenteista.

Sekä ensimmäisessä että kolmannessa kirjassa sivuhenkilöillä (taulukossa rivi ”muut”) *ut*- ja *quod*-konjunktioita esiintymistiheydet ovat selvästi keskimääräistä pienempiä. Tämä saattaa toki heijastella sitä, että sivuhenkilöiden puheessa on vähemmän sen tyyppisiä, näiden konjunktioita käyttöä vaativaa asioiden selittämistä, kuin päähenkilöiden esitelmätyyppisissä puheenvuoroissa. Vastaavaa eroa ei kuitenkaan voida havaita *cum*-konjunktioita kohdalla.

Sitä vastoin *cum*-konjunktioita suhteellisissa esiintymisfrekvensseissä on ensimmäisessä ja toisessa kirjassa huomattavia eroja pääpuhujien välillä. Ensimmäisessä kirjassa *cum* on Stololla suhteellisesti lähes kaksi kertaa yleisempi kuin Scrofalla, ja toisessa kirjassa Cossinuksella selvästi harvinaisempi kuin muilla puhujilla. Scrofan kohdalla voidaan myös todeta, ettei ainakaan alistuskonjunktioita frekvenssien perusteella ole selvää, onko hänen käyttämänsä kieli ensimmäisessä ja toisessa kirjassa samanlaista.

Varron omissa puheenvuoroissa ja kertojalla huomiota herättää *si*-konjunktioita suhteellinen harvinaisuus kaikissa kirjoissa ja kaikkiin muihin puhujiin verrattuna. Koko

aineistossa *si* esiintyy näissä jaksoissa yhteensä 15 kertaa, mikä on 0,36 prosenttia Varron kaikista saneista. Kaikilla puhujilla yhteensä *si* esiintyy 249 kertaa eli 0,70 prosenttia korpuksen saneiden kokonaismäärästä.

Edellä luvussa 5.2 on jonkin verran käsitelty *ut*-rakenteen monia erilaisia käyttötarkoituksia ja merkityksiä. Vastaavaa monikäyttöisyyttä on monilla muillakin yleisimmistä alistuskonjunktioista, joten pelkkien esiintymisfrekvenssien tarkastelu antaa vasta hyvin karkean kuvan erilaisten lauserakenteiden käytöstä. Jos kuitenkin huomattavia suhteellisia eroja havaitaan jo lemموjen frekvensseissä, voidaan perustellusti epäillä, että puhujien välillä on kielellisiä eroja, mutta näiden erojen laadusta ei voida päätellä vielä kovin paljoa. Sen sijaan frekvenssien samankaltaisuus ei välttämättä tarkoita, ettei kyseisten konjunktioiden käytössä voisi olla muunlaisia systemaattisia eroja. Tätä tarkastellaan vielä lähemmin *ut*-rakenteiden osalta seuraavassa luvussa.

Taulukko 3: Alistuskonjunktioiden esiintymät puhujittain

	määrä					prosenttiosuus				
	ut	quod	ne	cum	si	ut	quod	ne	cum	si
1. kirja										
Varro / kertoja	23	7	10	8	4	1,99	0,61	0,87	0,69	0,35
Scrofa	148	84	54	33	55	2,10	1,19	0,76	0,47	0,78
Stolo	101	55	30	41	43	2,07	1,13	0,62	0,84	0,88
Muut	19	13	9	10	16	1,38	0,94	0,65	0,73	1,16
<i>Yhteensä 1. kirja</i>	291	159	103	92	118	2,01	1,10	0,71	0,64	0,82
2. kirja										
Varro / kertoja	31	16	5	14	8	1,93	0,99	0,31	0,87	0,50
Scrofa	47	26	23	22	16	1,90	1,05	0,93	0,89	0,65
Cossinius	31	20	9	9	15	1,66	1,07	0,48	0,48	0,80
Atticus	39	24	35	16	19	1,69	1,04	1,52	0,69	0,82
Muut	48	29	27	29	17	1,80	1,09	1,01	1,09	0,64
<i>Yhteensä 2. kirja</i>	196	115	99	90	75	1,79	1,05	0,91	0,82	0,69
3. kirja										
Varro / kertoja	22	14	7	12	3	1,61	1,02	0,51	0,88	0,22
Merula	110	46	58	49	38	1,91	0,80	1,01	0,85	0,66
Appius	27	21	21	13	10	1,54	1,20	1,20	0,74	0,57
Muut	18	8	13	13	5	1,33	0,59	0,96	0,96	0,37
<i>Yhteensä 3. kirja</i>	177	89	99	87	56	1,73	0,87	0,97	0,85	0,55
Kaikki yhteensä	664	363	301	269	249	1,86	1,02	0,85	0,76	0,70

Rinnastuskonjunktiot

Rinnastuskonjunktiot eivät yhtä suoraan heijasta puhujien käyttämiä lauserakenteita kuin edellä käsitellyt alistuskonjunktiot, mutta ne voivat kertoa jotakin kielellisestä tyylistä. Selvästi yleisin rinnastuskonjunktio aineistossa on *et* yli tuhannella esiintymällä. Toisena tulee 359 esiintymällä *atque*, jota seuraavat *aut* ja enklittinen *-que* 257 ja 247 esiintymällään. Näiden määrät ja osuudet eri puhujilla on esitetty alla taulukossa 4. Rinnastuskonjunktioiden käyttö on kauttaaltaan, ja erityisen selvästi ensimmäisessä kirjassa, harvinaisempaa

sivuhenkilöillä kuin muilla puhujilla. Tämä on kuitenkin odotettavissa, sillä sivuhenkilöiden puheenvuorot ovat muutenkin lyhyempiä kuin muilla puhujilla.

Muihin puhujiin verrattuna Varron omissa jaksoissa *et* on yleisempi kun taas *atque* ja *-que* vastaavasti harvinaisempia. Tämä voisi olla viite siitä, että Varro on suosinut kielessä arkaaisia piirteitä erityisesti muiden puhujien suuhun pantuina. Tulkintaa vastaan puhuu kuitenkin todistusaineiston puute siitä, että *et* olisi jo tasavallan loppupuolella alkanut syrjäyttää *atque*- ja *-que*-konjunktioita.¹⁰⁵ Joka tapauksessa erot rinnastuskonjunktioiden suhteellisissa frekvensseissä Varron ja muiden puhujien välillä ovat selviä.

Taulukko 4: Rinnastuskonjunktioiden esiintymät puhujittain

	määrä				prosenttiosuus			
	et	atque	aut	-que	et	atque	aut	-que
1. kirja								
Varro / kertoja	34	10	4	12	2,95	0,87	0,35	1,04
Scrofa	207	81	63	54	2,93	1,15	0,89	0,76
Stolo	146	51	44	22	3,00	1,05	0,90	0,45
Muut	24	11	9	4	1,74	0,80	0,65	0,29
Yhteensä 1. kirja	411	153	120	92	2,84	1,06	0,83	0,64
2. kirja								
Varro / kertoja	72	9		5	4,47	0,56	0,00	0,31
Scrofa	71	20	12	20	2,86	0,81	0,48	0,81
Cossinius	59	25	8	9	3,16	1,34	0,43	0,48
Atticus	79	17	18	22	3,42	0,74	0,78	0,95
Muut	84	25	17	22	3,15	0,94	0,64	0,83
Yhteensä 2. kirja	365	96	55	78	3,34	0,88	0,50	0,71
3. kirja								
Varro / kertoja	45	12	8	8	3,29	0,88	0,58	0,58
Merula	148	59	54	43	2,57	1,03	0,94	0,75
Appius	56	22	17	13	3,19	1,25	0,97	0,74
Muut	29	17	3	13	2,15	1,26	0,22	0,96
Yhteensä 3. kirja	278	110	82	77	2,72	1,08	0,80	0,75
Kaikki yhteensä	1054	359	257	247	2,96	1,01	0,72	0,69

Prepositiot

Taulukko 5 esittää aineiston yleisimpien prepositioiden (*in*, *ad*, *ex*, *ab*, *de*, *cum*) määrät ja suhteelliset osuudet. Yleisesti ottaen prepositioiden käytössä ei ole kovin huomattavia eroja eri puhujien välillä. Näyttää kuitenkin siltä, että ensimmäisessä kirjassa Stolo ja toisessa kirjassa Cossinius käyttävät kaiken kaikkiaan muita puhujia enemmän prepositioita. Erityisesti mainittakoon, että ensimmäisessä kirjassa *cum*-preposition esiintymissä on Scrofan ja Stolon välillä yhtä voimakas ja samansuuntainen ero kuin edellä *cum*-konjunktion yhteydessä havaittu, joten mahdolliset jäsenysvirheet eivät voi olla ainakaan molempien erojen takana.

¹⁰⁵ Tällaisesta kehityksestä 1. vuosisadalta eaa. 4. vuosisadalle jaa. ks. Mitrović 2014, 102–203.

Taulukko 5: Prepositioiden esiintymät puhujittain

	määrä						prosenttiosuus					
	in	ad	ex	ab	de	cum	in	ad	ex	ab	de	cum
1. kirja												
Varro / kertoja	18	15	6	13	14	2	1,56	1,30	0,52	1,13	1,21	0,17
Scrofa	232	95	60	44	44	18	3,29	1,35	0,85	0,62	0,62	0,25
Stolo	197	65	50	28	30	19	4,04	1,33	1,03	0,57	0,62	0,39
Muut	40	26	6	12	10	2	2,90	1,89	0,44	0,87	0,73	0,15
<i>Yhteensä 1. kirja</i>	487	201	122	97	98	41	3,37	1,39	0,84	0,67	0,68	0,28
2. kirja												
Varro / kertoja	54	21	18	29	18	3	3,36	1,31	1,12	1,80	1,12	0,19
Scrofa	69	28	22	36	20	6	2,78	1,13	0,89	1,45	0,81	0,24
Cossinius	65	36	13	28	23	4	3,48	1,93	0,70	1,50	1,23	0,21
Atticus	53	39	15	24	21	8	2,30	1,69	0,65	1,04	0,91	0,35
Muut	76	40	22	29	17	9	2,85	1,50	0,83	1,09	0,64	0,34
<i>Yhteensä 2. kirja</i>	317	164	90	146	99	30	2,90	1,50	0,82	1,34	0,91	0,27
3. kirja												
Varro / kertoja	46	29	14	24	10	6	3,36	2,12	1,02	1,75	0,73	0,44
Merula	178	71	78	32	26	13	3,09	1,23	1,36	0,56	0,45	0,23
Appius	66	13	16	13	9	5	3,76	0,74	0,91	0,74	0,51	0,29
Muut	36	15	12	5	9	8	2,67	1,11	0,89	0,37	0,67	0,59
<i>Yhteensä 3. kirja</i>	326	128	120	74	54	32	3,19	1,25	1,17	0,72	0,53	0,31
Kaikki yhteensä	1130	493	332	317	251	103	3,17	1,38	0,93	0,89	0,70	0,29

Pronomininit

Pronominesita yli sata kertaa aineistossa esiintyvät relatiivipronomini *qui*, joka on myös yleisin lemma koko tekstissä, sekä demonstratiivipronomininit *is* ja *hic*. Näiden määrät ja osuudet on esitetty taulukossa 6. Etenkin *is*-pronominin ja *ire*-verbin taivutusmuotojen homonymiasta johtuvat virheet automaattisessa jäsennyksessä saattavat heikentää näiden tulosten luotettavuutta. Lisäksi on syytä huomata, että pronominit voivat toimia sekä itsenäisenä lausekkeina (kuten subjekteina ja objekteina) että attribuutteina. Tässä tarkastelussa käyttötapoja ei voida erottaa toisistaan, mutta näidenkin analysointi olisi periaatteessa mahdollista seuraavassa luvussa esitetyillä puupankkihauilla.

Pronominien esiintymien vertailu nostaa esiin muutamia mahdollisia eroja puhujien kielenkäytössä. Varron puheenvuoroissa ja kertojalla *is* on kaikissa kirjoissa suhteellisesti harvinaisempi ja *hic* vastaavasti yleisempi kuin muilla puhujilla. Toisessa kirjassa Cossinus ja kolmannessa Appius käyttävät näitä yleisimpiä pronomineja ylipäättään enemmän kuin muut puhujat; molemmat suosivat erityisesti relatiivipronominia, Appius lisäksi *hic*-pronominia. Sivuhenkilöillä *is* on suhteellisen harvainen sekä ensimmäisessä että kolmannessa kirjassa. Nämä erot saattavat liittyä puheenvuorojen aiheisiin, sillä

pronomineilla on erilaiset viittauskentät. Erityisesti henkilöillä, joilla on vain harvoja puheenvuoroja, yksittäinen aihe saattaa näkyä selvästi frekvenssitilastossa.

Taulukko 6: Pronominien esiintymät puhujittain

	määrä			prosenttiosuus		
	qui	is	hic	qui	is	hic
1. kirja						
Varro / kertoja	51	21	14	4,42	1,82	1,21
Scrofa	263	144	72	3,73	2,04	1,02
Stolo	197	121	25	4,04	2,48	0,51
Muut	46	13	15	3,34	0,94	1,09
<i>Yhteensä 1. kirja</i>	557	299	126	3,85	2,07	0,87
2. kirja						
Varro / kertoja	58	22	21	3,60	1,37	1,31
Scrofa	89	41	19	3,59	1,65	0,77
Cossinius	80	37	18	4,28	1,98	0,96
Atticus	64	47	17	2,77	2,04	0,74
Muut	72	48	32	2,70	1,80	1,20
<i>Yhteensä 2. kirja</i>	363	195	107	3,32	1,78	0,98
3. kirja						
Varro / kertoja	42	17	18	3,07	1,24	1,31
Merula	206	132	64	3,58	2,29	1,11
Appius	69	34	35	3,93	1,94	2,00
Muut	35	21	12	2,59	1,56	0,89
<i>Yhteensä 3. kirja</i>	352	204	129	3,44	1,99	1,26
Kaikki yhteensä	1272	698	362	3,57	1,96	1,02

Apuverbi ja kieltosana

Apuverbin *esse* ja kieltosanan *non* määrät ja osuudet on esitetty taulukossa 7. Taulukon oikeanpuoleisimmassa sarakkeessa on lisäksi apuverbin esiintymien määrä suhteessa kieltosanan esiintymien määrään. Apuverbin ja kopulan poisjätto on yksi Varron kielelle tyypillisistä piirteistä, joita on esitelty edellä luvussa 3.2, joten sen esiintymien tarkastelu on erityisen perusteltua. Kieltosana *non* kuuluu aineiston kymmenen yleisimmän lemman joukkoon, ja lisäksi on mahdollista, että kieltosana voi vaikuttaa apuverbin (ja muidenkin verbien) esiintymistiheyteen, koska sen yksi tyypillinen esiintymä on rinnasteisissa lauseissa (*non ... sed ... tms.*), joilla taas usein on yhteinen predikaatti.

Apuverbin tai kopulan esiintymiä on Scrofalla sekä ensimmäisessä että toisessa kirjassa suhteellisesti enemmän verrattuna muihin puhujiin, erityisesti Varroon sekä ensimmäisessä kirjassa Stoloon, toisessa kirjassa ryhmään ”muut” (Lucienus, Vaccius ja Murrius). Varron muita puhujia harvemmat *esse*-verbin esiintymät näyttävät kuitenkin rajoittuvan kahteen ensimmäiseen kirjaan. Kolmannessa kirjassa nimittäin Varron ja kertojan osuuksissa *esse* on huomattavasti yleisempi kuin kenelläkään muulla puhujalla missään kolmesta kirjasta.

Varron ja Scrofan jaksoissa kieltosanan suhteellinen yleisyys näyttää edellä käsiteltyjen apuverbin esiintymien peilikuvalta. Verbin *esse* keskimääräistä suurempi osuus (Scrofa 1. ja 2. kirja, Varro 3. kirja) esiintyy samanaikaisesti kieltosanan keskimääräistä pienemmän prosenttiosuuden kanssa, ja päinvastoin *essen* pienempi osuus (Varro 1. ja 2. kirja, Appius ja ryhmä muut 3. kirja) yhdessä kieltosanan keskimääräistä suuremman osuuden kanssa. Vastaavaa ei voida kovin selvästi havaita, jos verrataan kolmen kirjan esiintymäosuuksia keskenään: *sum-* ja *non-*esiintymien keskinäinen suhde on kaikissa kirjoissa suunnilleen sama. Esimerkiksi toisessa kirjassa kummatkin osuudet ovat hieman koko aineiston keskiarvoja pienemmät. Mahdollinen ero kielenkäytössä olisi siis puhujien välinen.

Taulukko 7: Apuverbin ja kieltosanan esiintymät puhujittain

	määrä		prosenttiosuus		apuverbin ja kieltosanan suhde [sum/non]
	sum	non	sum	non	
1. kirja					
Varro / kertoja	29	20	2,52	1,73	1,46
Scrofa	237	87	3,36	1,23	2,73
Stolo	130	58	2,67	1,19	2,24
Muut	42	21	3,05	1,52	2,01
Yhteensä 1. kirja	438	186	3,03	1,29	2,35
2. kirja					
Varro / kertoja	40	34	2,49	2,11	1,18
Scrofa	86	23	3,47	0,93	3,73
Cossinius	56	20	3,00	1,07	2,80
Atticus	67	26	2,90	1,13	2,57
Muut	67	24	2,52	0,90	2,80
Yhteensä 2. kirja	316	127	2,89	1,16	2,49
3. kirja					
Varro / kertoja	66	14	4,82	1,02	4,73
Merula	180	54	3,13	0,94	3,33
Appius	49	28	2,79	1,60	1,74
Muut	31	22	2,30	1,63	1,41
Yhteensä 3. kirja	326	118	3,19	1,15	2,77
Kaikki yhteensä	1080	431	3,03	1,21	2,50

6.3 Puupankki

Seuraavaksi käsittelen puupankin hyödyntämistä aineiston rakenteiden analysoinnissa. Arvioin ensin automaattisesti jäsennetyn aineiston luotettavuutta ja käyttökelpoisuutta. Havainnollistan sitten puupankkianalyysiä esimerkeillä, jotka kytkeytyvät luvussa 3.2 käsiteltyihin Varron kielen erityispiirteisiin sekä luvussa 5.2 esiteltyyn *ut*-rakenteiden moninaisuuteen. Tavoitteena on ensisijaisesti selvittää, kuinka menetelmä soveltuu käsillä olevan tutkimuskysymyksen käsittelyyn.

Automaattinen jäsennys

Koko aineisto on jäsennetty automaattisesti UDPipe-työkalulla kaikilla kolmella latinan kielelle saatavilla olevalla mallia. Eri mallien tulosten luotettavuuden arviointia varten olen verrannut mallien tuottamia tuloksia manuaalisesti annotoituun otteeseen ensimmäisen kirjan alusta, joka käsittää ilman välimerkkejä 2 360 sanetta. Malleja ja annotoituja otteita on käsitelty edellä luvussa 4.3. Vertailun tulokset on esitetty sanaluokittain taulukossa 8.

Sanaluokasta riippuen kaikki mallit tunnistavat lemmän 70–100 prosentissa otteen saneista. Etenkin verbien, substantiivien ja adjektiivien lemموjen tunnistamisesta Perseus- ja PROIEL-mallit suoriutuvat ITTB-mallia paremmin. Dependenssikieliopin mukaisen oikean pääsanatun tunnistamisessa automaattisen jäsennys onnistuu vaihtelevasti eri sanaluokille eri malleilla. Kaikki mallit tunnistavat alistuskonjunktioiden ja adpositioiden pääsanatun kohtuullisella 73–92 prosentin tarkkuudella. Rinnastuskonjunktioiden kohdalla vain ITTB pääsee lähellekään kohtuullista onnistumisprosenttia. Muiden sanaluokkien osalta ITTB-malli kuitenkin suoriutuu kahta muuta mallia selvästi heikommin.

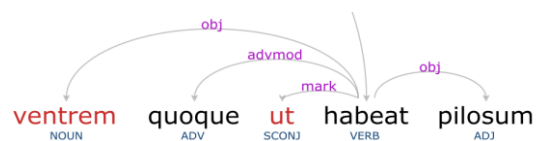
Rinnastuskonjunktioiden pääsanatun tunnistusongelmat eivät kuitenkaan luultavasti ole lauserakenteen analyysin onnistumisen kannalta kriittisiä. Todennäköinen selitys huonolle onnistusprosentille on, että PROIEL- ja Perseus-mallit ovat liittäneet konjunktion rinnasteisista elementeistä jollekin muulle kuin ensimmäiselle. Tällainen virhe on luonteeltaan tekninen, johtunee puupankkien aiemmissa versioissa käytetyistä vanhemmista annotointikäytännöistä, eikä oleellisesti muuta lauserakennetta sisällöllisesti. Koska Perseus-malli suoriutuu kaikkien sanaluokkien kohdalla hieman PROIEL-mallia paremmin, käytän tässä analyysissä jatkossa Perseus-mallin tuottamaa, koko aineistosta automaattisesti jäsennettyä puupankkia.

Taulukko 8: Automaattisen jäsennyksen onnistumisprosentit - UDPipe ja eri mallit

Sanaluokka	Saneita	lemma			pääsana		
		ITTB	PROIEL	Perseus	ITTB	PROIEL	Perseus
Verbi	528	75 %	89 %	88 %	52 %	59 %	62 %
Substantiivi	713	73 %	85 %	86 %	56 %	57 %	64 %
Pronomini	242	81 %	85 %	85 %	61 %	65 %	69 %
Adjektiivi, numeraali	216	75 %	79 %	85 %	61 %	61 %	65 %
Adverbi	233	86 %	86 %	86 %	65 %	61 %	72 %
Rinnastuskonjunktio	132	97 %	100 %	100 %	67 %	11 %	15 %
Alistuskonjunktio	126	95 %	92 %	94 %	73 %	75 %	80 %
Adpositio	170	99 %	99 %	100 %	88 %	88 %	92 %
Yhteensä	2360	80 %	88 %	88 %	61 %	60 %	65 %

Anastrofi

Olen analysoinut Luvussa 4.3 mainitulla Tundra-sovelluksella Varron kielelle tyypillisten anastrofien esiintymistä puupankissa. Tarkastelu on rajattu aineiston neljään yleisimpään alistuskonjunktioon *ut*, *si*, *quod* ja *cum*. Anastrofien tunnistamista varten puupankista haetaan näiden konjunktoiden sellaisia *sisarelementtejä*, jotka tulevat tekstissä ennen tutkittavana olevaa konjunktia. Sisarelementillä tarkoitetaan tässä sellaisia lausekkeita, joilla on yhteinen pääsana. Tulokset eivät kerro suoraan anastrofien määrää, koska haku listaa kaikki konjunktia edeltävät sisarelementit, joita voi olla hyvinkin monia. Asian havainnollistamiseksi tarkastellaan lausetta *ventrem quoque ut habeat pilosum*:



Huomaamme, että automaattinen jäsennin on erehtynyt adjektiivin *pilosum* pääsanasta (pitäisi tietenkin olla attribuutti *ventrem*-sanalle), mikä ehkä kerto tällaisten rakenteiden tietoteknisille apuvälineille asettamista haasteista. Tämä ei onneksi vaikuta nyt käsillä olevaan kysymykseen, koska olemme kiinnostuneita vain *ut*-sanaa edeltävistä sisarelementeistä. Tässä tapauksessa sekä *ventrem* että *quoque* tulevat mukaan laskuihin.

Taulukon 9 vasemmanpuoleisessa osassa on tarkasteltavia konjunktioita edeltävien sisarelementtien määrät puhujittain. Oikeanpuoleisessa osassa nämä lukumäärät on suhteutettu kyseisten konjunktoiden esiintymien määrään ko. puhujalla. Luvut eivät siis kerro varsinaista anastrofien määrää vaan sen, kuinka monta sisarelementtiä on konjunktoiden edellä anastrofien seurauksena. Suurin osa puhujakohtaisista luvuista on sen verran pieniä, ettei niiden perusteella voi tehdä vahvoja tilastollisia päätelmiä. Ainoat huomiota herättävät seikat ovat, että Scrofalla *ut*-konjunktion edellä näyttäisi olevan sisarelementtejä huomattavasti vähemmän kuin muilla puhujilla ja Varron omissa jaksoissa konjunktista riippumatta kaiken kaikkiaan enemmän kuin muilla.

Taulukko 9: Konjuktioiden *ut*, *si*, *quod* ja *cum* anastrofi puhujittain

	<i>edeltävät sisarelementit</i>					<i>edeltävät sisarelementit / konjunktio, prosenttia</i>				
	<i>ut</i>	<i>si</i>	<i>quod</i>	<i>cum</i>	<i>Yhteensä</i>	<i>ut</i>	<i>si</i>	<i>quod</i>	<i>cum</i>	<i>Yhteensä</i>
Varro / kertoja	15	2	9	17	43	19,7	13,3	27,3	37,8	25,4
Scrofa	33	10	16	17	76	16,8	14,7	16,2	21,3	17,1
Merula	20	5	4	11	40	18,2	13,5	9,5	17,7	15,9
Stolo	23	5	4	13	45	22,8	11,6	8,3	21,7	17,9
Appius	10	3	3	3	19	37,0	30,0	11,5	16,7	23,5
Atticus	10	3	6	3	22	25,6	15,8	22,2	12,5	20,2
Cossinius	9	1	2	1	13	29,0	6,7	9,1	7,7	16,0
Muut	21	6	8	9	44	24,4	15,8	18,6	12,7	18,5
Yhteensä	141	35	52	74	302	21,1	14,3	15,3	19,8	18,6

Samoja sisarelementtejä on tarkasteltu taulukossa 10 sen mukaan, mikä tällaisen edeltävän elementin dependenssirelaatio pääsanaan on. Lauseenjäsenistä subjekti ja objekti on esitetty erikseen. Sarakkeen ”adv. ym.” luvut koostuvat suurimmaksi osaksi tavanomaisista yhden sanan adverbimääreistä mutta sisältävät myös kokonaisia adverbiaalilauseita sekä joitakin hajanaisia muita nominien määreitä kuin subjekti tai objekti. Taulukon oikean puolen prosenttiosuudet kuvaavat erityyppisten anastrofisten sisarelementtien suhteellisia osuuksia.

Jakaumat ovat eri puhujilla melko saman tyyppisiä, mutta Scrofan ja Stolon välillä näyttäisi tässäkin olevan pieni ero. Scrofalla adverbiaalilausekkeet konjunktioiden edellä ovat muita lauseenjäseniä yleisempiä, Stololla taas harvinaisempia.

Taulukko 10: Anastrofi – edeltävän lausekkeen tyyppi

	<i>määrät</i>				<i>prosenttiosuudet</i>		
	<i>adv. ym.</i>	<i>subjekti</i>	<i>objekti</i>	<i>Yhteensä</i>	<i>adv. ym.</i>	<i>subjekti</i>	<i>objekti</i>
Varro / kertoja	27	8	8	43	62,8	18,6	18,6
Scrofa	54	11	11	76	71,1	14,5	14,5
Merula	26	6	8	40	65,0	15,0	20,0
Stolo	26	9	10	45	57,8	20,0	22,2
Appius	11	4	4	19	57,9	21,1	21,1
Atticus	10	7	5	22	45,5	31,8	22,7
Cossinius	7	4	2	13	53,8	30,8	15,4
Muut	28	8	8	44	63,6	18,2	18,2
Yhteensä	189	57	56	302	62,6	18,9	18,5

Ut-rakenteiden ominaisuuksia

Luvussa 5.2 on selvitetty *ut*-rakenteiden annotointiin ja analyysiin liittyviä kysymyksiä. Ut-sanat esiintymistä eri puhujilla on tutkittu edellä luvussa 6.2. Saadakseni lisävalaistusta näiden rakenteiden käytöstä tutkittavassa aineistossa eri puhujilla olen analysoinut Tundra-sovelluksella puupankin *ut*-lauseiden ja -lausekkeiden pääsanojen sanaluokkia. Lisäksi olen tarkastellut moduksia niissä *ut*-lauseissa, joissa *ut* liittyy joko suoraan finiittiverbiin tai partisiippi- tai predikaatiivirakenteen välityksellä apuverbiin tai kopulaan.

Alustava Tundra-haku osoittaa, että *ut* saa puupankissa pääsanakseen verbejä, substantiiveja ja adjektiiveja. Satunnaisesti pääsanana on myös muita sanaluokkia, mutta nämä jätän todennäköisinä jäsennysvirheinä pois tästä tarkastelusta. Esitän tulokset jaettuna nomineihin ja verbeihin. Alustava haku ei kuitenkaan anna oikeaa kuvaa pääsanojen jakaumasta, mikä johtuu UD:n konventiosta jäsentää predikatiivilausekkeessa pääsanaksi predikatiivi, josta kopula riippuu. Haluttujen tulosten saamiseksi haku on tehtävä useammassa osassa:

- Ensin etsitään puupankista erikseen *ut*-rakenteiden pääsanoina olevat verbit (Tundra-haku: [pos="VERB"] > [word="ut"]) sekä pääsanat, jotka eivät ole verbejä ([pos!="VERB"] > [word="ut"]).
- Tämän jälkeen haetaan jälkimmäisistä nomini-pääsanoista sellaiset, joilla on sisarelementtinä kopula. ([pos!="VERB"] > [word="ut"] \$ [edge="cop"]).
- Vähentämällä jälkimmäisessä haussa saadut tulokset nomini-pääsanojen määrästä ja vastaavasti lisäämällä ne verbi-pääsanojen lukumäärään saadaan erilaisia *ut*-rakenteita syntaktisesti paremmin kuvaava jaottelu.

Näin lasketut *ut*-rakenteiden lukumäärät sekä niiden suhteelliset osuudet prosentteina on esitetty taulukossa 11. Noin 80 prosentissa aineiston tapauksia *ut*-rakenteen välittömänä pääsanana on verbi ja noin 20 prosentissa nomini. Puhujien välillä eroja on siten, että Scrofalla ja Stollolla pääsanana on muita puhujia useammin nomini, kun taas Varron ja Merulan jaksoissa nomini pääsanana on harvinainen. Muiden puhujien kohdalla esiintymiä on johtopäätösten tekemiseksi vähän.

Taulukko 11: Ut-rakenteiden pääsanat

	määrät			prosenttiosuudet	
	Nomini	Verbi	Yhteensä	Nomini	Verbi
Varro / kertoja	11	64	75	14,7	85,3
Scrofa	50	142	192	26,0	74,0
Merula	15	94	109	13,8	86,2
Stolo	23	77	100	23,0	77,0
Appius	5	22	27	18,5	81,5
Atticus	4	34	38	10,5	89,5
Cossinius	6	24	30	20,0	80,0
Muut	16	68	84	19,0	81,0
Yhteensä	130	525	655	19,8	80,2

Selvitettäessä *ut*-lauseiden finiittiverbien moduksia on samaan tapaan kuin edellä tehtävä useampia hakuja. Kun pääsanana on finiittiverbi, modus saadaan suoraan puupankkihaun tuloksena olevasta taulukosta. Vastaavasti saadaan kopulan modus edellä mainittuihin predikatiivirakenteisiin. Näiden lisäksi on liittomuotojen modus selvitettävä kuten

predikatiivirakenteille, hakemalla partisiippien sisarelementteinä olevat apuverbit. Näiden kolmen ryhmän yhteenlasketut, indikatiivissa ja konjunktiiivissa olevien *ut*-lauseiden määrät on esitetty taulukossa 12.¹⁰⁶

Noin 30 prosenttia *ut*-lauseista on indikatiivissa ja loput 70 prosenttia konjunktiiivissa. Varron omia jaksoja lukuun ottamatta eniten tekstissä esiintyvät puhujat Scrofa, Merula ja Stolo ovat sangen lähellä tätä suhdelukua. Varrolla sekä Appiuksella ja Cossiniuksella indikatiivi näyttää huomattavan yleiseltä. Tämä saattaa olla merkki esimerkiksi *ut*-sanan adverbien omaisesta käytöstä tai temporaalisten *ut*-lauseiden yleisyydestä.

Taulukko 12: Verbien modukset ut-lauseissa

	<i>määrät</i>			<i>prosenttiosuudet</i>	
	Indikatiivi	Konjunktiiivi	<i>Yhteensä</i>	Indikatiivi	Konjunktiiivi
Varro / kertoja	24	39	63	38,1	61,9
Scrofa	38	98	136	27,9	72,1
Merula	24	66	90	26,7	73,3
Stolo	21	52	73	28,8	71,2
Appius	9	13	22	40,9	59,1
Atticus	6	28	34	17,6	82,4
Cossinius	11	12	23	47,8	52,2
Muut	18	47	65	27,7	72,3
Yhteensä	151	355	506	29,8	70,2

¹⁰⁶ Laskelmasta on jätetty pois aineistossa olevat muutama gerundiivi sekä yksi imperatiivi.

7 YHTEENVETO JA JOHTOPÄÄTÖKSET

Olen pyrkinyt selvittämään, onko Varron *Res Rusticae* -teoksen dialogimuotoisessa tekstissä havaittavissa kielellisiä eroja puhujien välillä. Analyysin pohjana olen käyttänyt teoksen kaikki kolme kirjaa kattavaa digitaalista tekstikorpusta. Olen jäsentänyt ja annotoinut aineistoa sekä manuaalisesti että automaattisen jäsentimen avulla. Jäsennyksen tuloksena olevasta puupankista olen tarkastellut valikoituja kielellisiä ominaisuuksia eri puhujien teksteissä.

Digitaalinen aineisto tekee mahdolliseksi määrällisen analyysin useista erilaisista kielellisistä piirteistä melko vähäisellä lisätyöllä sen jälkeen, kun aineisto on kunnolla valmisteltu. Käytetyt menetelmät soveltuvatkin hyvin tunnistamaan aineistosta mahdollisesti huomionarvoisia seikkoja. Erilaisten leksikaalisten ja syntaktisten piirteiden takana voi kuitenkin olla moninaisia syitä ja kielellisiä ilmiöitä, joten useimmissa tapauksissa menetelmien tuottamat havainnot toimivat lähinnä hyvinä vihjeinä lähemmän tutkimuksen kohteiksi. Tässä työssä esimerkinomaisesti tarkastelluissa funktiosanojen frekvensseissä, alustuskonjunktioiden anastrofissa ja ut-rakenteiden ominaisuuksissa havaitaan puhujien välillä sellaisia eroja, että niiden tutkiminen perusteellisemmin vaikuttaa mielekkäältä.

Jatkotarkastelun kannalta tulosten luotettavuutta kasvattaisi digitaalisen aineiston parantaminen. Tätä varten automaattisen jäsennyksen tulokset olisi tarkistettava nykyistä suuremmasta osasta tekstiä ja mieluiten koko korpuksesta. Tämä on tehtävä viime kädessä manuaalisesti, mutta digitaaliset työkalut tarjoavat myös mahdollisuuksia jäsennysmallien ”kouluttamiseen”, mikä voi nopeuttaa puupankin laadun parantamista.

Jo aiemmassa tutkimuksessa on havaittu eräitä Varron kielelle ominaisia piirteitä, joiden esiintymistä erityisesti dialogin eri puhujille voidaan tarkastella digitaalisesta aineistosta myös huomattavasti monipuolisemmin kuin mitä tämän työn puitteissa on ollut mahdollista esittää. Lisäksi tämänkaltaisen aineiston yleiskuvan tarkastelu parhaimmillaan tuottaa uusia havaintoja sellaisista erityispiirteistä, joihin ei välttämättä tulisi kiinnittäneeksi huomiota pelkästään tekstiä lukemalla.

Ediot

- Flach, D. (1996). *Gespräche über die Landwirtschaft - Marcus Terentius Varro ; herausgegeben, übersetzt und erläutert von Dieter Flach*. Wissenschaftliche Buchgesellschaft.
- Guiraud, C. (1985). *Économie rurale / T. 2, Livre II / texte établi, trad. et commenté par Charles Guiraud*. Les Belles lettres.
- Guiraud, C. (1997). *Économie rurale / T. 3, Livre III / texte établi, trad. et commenté par Charles Guiraud*. Les Belles lettres.
- Heurgon, J. (1978). *Économie rurale / Varron ; texte établi, traduit et commenté par Jacques Heurgon*. Belles Lettres,.
- Hooper, W. D., Ash, H. B., Goetz, G. (1934). *On agriculture*. Harvard University Press ; W. Heinemann.

Tutkimuskirjallisuus

- Bamman, D., & Crane, G. (2006). The design and use of a Latin dependency treebank. *Proceedings of the Fifth Workshop on Treebanks and Linguistic Theories (TLT2006)*, 67–78.
- Bolkestein, A. M. (1977). Part II: The Differences between Free and Obligatory ut-Clauses. *Glotta*, 55(3/4), 231–250. JSTOR.
- Bresnan, J. (2001). *Lexical-functional syntax*. Blackwell.
- Buchholz, S., & Marsi, E. (2006). CoNLL-X Shared Task on Multilingual Dependency Parsing. *Proceedings of the Tenth Conference on Computational Natural Language Learning (CoNLL-X)*, 149–164. <https://www.aclweb.org/anthology/W06-2920>
- Burns, P. J. (2019). Building a text analysis pipeline for Classical languages. *Digital Classical Philology: Ancient Greek and Latin in the Digital Revolution*, 10, 159.
- Butterfield, D. J. (Ed.). (2015). *Varro Varius: the polymath of the Roman world*. Cambridge Philological Society.
- Castrén, P., & Pietilä-Castrén, L. (2000). *Antiikin käsikirja*. Otava.
- Cèbe, J.-P. (1972). *Varron, Satires ménippées: Édition, traduction et commentaire*. Ecole française de Rome.
- Cecchini, F. M., Passarotti, M., Marongiu, P., & Zeman, D. (2018). *Challenges in Converting the Index Thomisticus Treebank into Universal Dependencies*. <https://doi.org/10.5281/ZENODO.2552203>
- De Marneffe, M.-C., Dozat, T., Silveira, N., Haverinen, K., Ginter, F., Nivre, J., & Manning, C. D. (2014). Universal Stanford dependencies: A cross-linguistic typology. *LREC*, 14, 4585–4592.
- Della Corte, F. (1954). *Varrone: Il terzo gran lume romano*. Pubblicazioni dell'istituto universitario di magisterio.
- Diederich, S. (2007). *Römische Agrarhandbücher zwischen Fachwissenschaft, Literatur und Ideologie*. Walter de Gruyter.

- Gentili, G. (1903). *De Varronis in libri rerum rusticarum auctoribus*. Seeber.
- Green, C. M. C. (1997). Free as a Bird: Varro de re Rustica 3. *The American Journal of Philology*, 118(3), 427–448.
- Hajič, J., Panevová, J., Hajičová, E., Sgall, P., Pajas, P., Štěpánek, J., Havelka, J., Mikulová, M., Žabokrtský, Z., Ševčíková-Razímová, M., & Urešová, Z. (2006). *Prague Dependency Treebank 2.0*.
- Harrington, J. M. (n.d.). *Tufts University Treebanked Commentaries*. The Perseids Project. Retrieved 1 August 2020, from <https://perseids-publications.github.io/harrington-trees/>
- Haug, D. T., & Jøhndal, M. (2008). Creating a parallel treebank of the old Indo-European Bible translations. *Proceedings of the Second Workshop on Language Technology for Cultural Heritage Data (LaTeCH 2008)*, 27–34.
- Heinecke, J. (2019). ConlluEditor: A fully graphical editor for Universal dependencies treebank files. *Proceedings of the Third Workshop on Universal Dependencies (UDW, SyntaxFest 2019)*, 87–93. <https://doi.org/10.18653/v1/W19-8010>
- Heisterhagen, R. (1952). *Die literarische Form der Rerum rusticarum libri Varros*. Philipps-Universität Marburg.
- Horsfall, N. (1982). Prose and mime. *Cambridge History of Classical Literature*, 2(part 2), 112–20.
- Karlsson, F. (2006). *Yleinen kielitiede: Uudistettu laitos*. Yliopistopaino.
- Kenney, E. J. (1982). Books and readers in the Roman world. *The Cambridge History of Classical Literature*, 2, 3–32.
- Korkiakangas, T. (2021). Late Latin Charter Treebank: Contents and annotation. *Corpora*, 2021(16).
- Koskeniemi, K. (2002). *Kieliteknologian termistö*. <http://www.ling.helsinki.fi/kit/term/intro.shtml>
- Kronenberg, L. (2009). *Allegories of farming of Greece and Rome*. Cambridge University Press.
- Laughton, E. (1960). Observations on the Style of Varro. *The Classical Quarterly*, 10(1–2), 1–28. <https://doi.org/10.1017/S0009838800024332>
- Martens, S. (2013). TüNDRA: A web application for treebank search and visualization. *The Twelfth Workshop on Treebanks and Linguistic Theories (TLT12)*, 133.
- Martin, René. (1971). *Recherches sur les agronomes latins et leurs conceptions économiques et sociales*.
- Martin, RENÉ. (1995). Ars an quid aliud? La conception Varronienne de l'agriculture. *Revue Des Études Latines*, 73, 80–91.
- Mitrović, M. (2014). *Morphosyntactic atoms of propositional logic: A philo-logical programme*. Diss. University of Cambridge.
- Nelstuen, G. A. (2015). *Varro the agronomist: Political philosophy, satire, and agriculture in the late Republic*. The Ohio State University Press.
- Nivre, J., De Marneffe, M.-C., Ginter, F., Goldberg, Y., Hajič, J., Manning, C. D., McDonald, R., Petrov, S., Pyysalo, S., & Silveira, N. (2016). Universal dependencies v1: A

- multilingual treebank collection. *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, 1659–1666.
- Passarotti, M. C. (2016). *How far is Stanford from Prague (and vice versa)? comparing two dependency-based annotation schemes by network analysis*.
- Passarotti, M. C. (2011). Language resources. The state of the art of Latin and the Index Thomisticus treebank project. *Deuxième Colloque International ALIENTO*, 301–320.
- Petrov, S., Das, D., & McDonald, R. (2011). A universal part-of-speech tagset. *ArXiv Preprint ArXiv:1104.2086*.
- Purcell, N. (1995). The Roman villa and the landscape of production. In *Urban Society in Roman Italy* (pp. 167–194). UCL Press.
- Riposati, B. (1975). *M. Terenzio Varrone, l'uomo e lo scrittore*. Istituto di studi romani.
- Rodgers, R. H. (1978). Varro and Virgil in the Geoponica. *Greek, Roman, and Byzantine Studies*, 19(3), 277–285.
- Rodgers, R. H. (2015). *A New Text of De Re Rustica*. Varro Varius: The Polymath of the Roman World. Cambridge: Cambridge
- Rosa, R., Masek, J., Marecek, D., Popel, M., Zeman, D., & Zabokrtský, Z. (2014). HamleDT 2.0: Thirty Dependency Treebanks Stanfordized. *LREC*, 2334–2341.
- Sgall, P., Hajičová, E., Panevová, J., & Mey, J. (1986). *The meaning of the sentence in its semantic and pragmatic aspects*. D. Reidel ; Sold and distributed in the U.S.A. and Canada by Kluwer Academic.
- Skydsgaard, J. E. (1968). *Varro the scholar: Studies in the first book of Varro's De re rustica*. E. Munksgaard.
- Straka, M., Hajič, J., & Strakova, J. (n.d.). *UDPipe: Trainable Pipeline for Processing CoNLL-U Files Performing Tokenization, Morphological Analysis, POS Tagging and Parsing*. 8.
- Straka, M., & Strakova, J. (2017). Tokenizing, POS Tagging, Lemmatizing and Parsing UD 2.0 with UDPipe. 88–99. <https://doi.org/10.18653/v1/K17-3009>
- Trevizam, M. (2011). A atenção de Varro às palavras no livro III do 'De re rustica'. *Humanitas*, 63, 355–371.
- Trevizam, M. (2012). A face 'gramatical' de Varrão em De Re Rustica II. *Aletria: Revista de Estudos de Literatura*, 19(3).
- Trevizam, M. (2013). Linguagem e gênero na literatura agrária latina: Catão, Varrão e Virgílio. *Classica - Revista Brasileira de Estudos Clássicos*, 20(1), 7–18. <https://doi.org/10.24277/classica.v20i1.130>
- Trevizam, M. (2014). Distintos tratamentos lexicais e textuais dos plantéis pelo Virgílio de Geórgicas III e por Varrão (De re rustica II). *PhaoS-Revista de Estudos Clássicos*, 12(12).
- White, K. D. (1973). Roman agricultural writers I: Varro and his predecessors. *Aufstieg Und Niedergang Der Römischen Welt I*, 4, 439–497.
- Zeman, D. (2008). Reusable Tagset Conversion Using Tagset Drivers. *LREC*, 2008, 28–30.

LITTEET

Lüite 1: Näytteitä aineiston käsittelyssä käytettävistä tiedostoista ja työkaluista

Conll-U

newpar id = rust.1.2.2

sent_id = 31

text = Nos vero , inquit Agrius , ut arbitror , eadem causa quae te , rogatio aeditumi .

1	Nos	nos	PRON	_	Case=Acc Gender=Masc Number=Plur	11	nsubj	_	Speaker=Agrius
2	vero	verus	ADJ	_	Case=Abl Gender=Neut Number=Sing	1	advmod	_	Speaker=Agrius
3	,	,	PUNCT	_	_	4	punct	_	_
4	inquit	inquam	VERB	_	Mood=Ind Number=Sing Person=3 Tense=Pres VerbForm=Fin Voice=Act	1	parataxis	_	Speaker=Agrius
5	Agrius	Agrius	PROPN	_	Case=Nom Gender=Masc Number=Sing	4	nsubj	_	Speaker=Agrius
6	,	,	PUNCT	_	_	7	punct	_	_
7	ut	ut	SCONJ	_	_	8	mark	_	Speaker=Agrius
8	arbitror	arbitror	VERB	_	Mood=Ind Number=Sing Person=1 Tense=Pres VerbForm=Fin Voice=Pass	1	advcl	_	Speaker=Agrius
9	,	,	PUNCT	_	_	10	punct	_	_
10	eadem	idem	PRON	_	Case=Nom Gender=Fem Number=Sing	11	nmod	_	Speaker=Agrius
11	causa	causa	NOUN	_	Case=Nom Gender=Fem Number=Sing	0	root	_	Speaker=Agrius
12	quae	quis	NOUN	_	Case=Nom Gender=Fem Number=Plur	13	nmod	_	Speaker=Agrius
13	te	tu	PRON	_	Case=Abl Gender=Masc Number=Sing	11	obl	_	Speaker=Agrius
14	,	,	PUNCT	_	_	15	punct	_	_
15	rogatio	rogatio	NOUN	_	Case=Nom Gender=Fem Number=Sing	11	appos	_	Speaker=Agrius
16	aeditumi	aeditumus	NOUN	_	Case=Gen Gender=Masc Number=Sing	15	nmod	_	Speaker=Agrius
17	.	.	PUNCT	_	_	4	punct	_	_

Arethusa xml:

```
<sentence id="4" document_id="" subdoc="" span="">
  <word id="1" form="Nos" lemma="nos" postag="p-p---ma-" relation="OBJ" head="18"/>
  <word id="2" form="vero" lemma="verus" postag="a-s---nb-" relation="AuxY" head="18"/>
  <word id="3" form="," lemma="punc1" postag="u-----" relation="AuxX" head="18"/>
  <word id="4" form="inquit" lemma="inquam" postag="v3spia---" relation="ExD" head="18"/>
  <word id="5" form="Agrius" lemma="agrius1" postag="a-s---mn-" relation="SBJ" head="4"/>
  <word id="6" form="," lemma="punc1" postag="u-----" relation="AuxX" head="18"/>
  <word id="7" form="ut" lemma="ut" postag="c-----" relation="AuxC" head="18"/>
  <word id="8" form="arbitror" lemma="arbitror" postag="v1spip---" relation="ADV" head="7"/>
  <word id="9" form="," lemma="punc1" postag="u-----" relation="AuxX" head="18"/>
  <word id="10" form="eadem" lemma="idem" postag="p-s---fn-" relation="ATR" head="11"/>
  <word id="11" form="causa" lemma="causa" postag="n-s---fn-" relation="SBJ_AP" head="14"/>
  <word id="12" form="quae" lemma="quis1" postag="n-p---fn-" relation="SBJ" head="19"/>
  <word id="13" form="te" lemma="tu" postag="p-s---mb-" relation="OBJ" head="19"/>
  <word id="14" form="," lemma="punc1" postag="u-----" relation="APOS" head="18"/>
  <word id="15" form="rogatio" lemma="rogatio" postag="n-s---fn-" relation="SBJ_AP" head="14"/>
  <word id="16" form="aeditumi" lemma="aeditumus" postag="n-s---mg-" relation="ATR" head="15"/>
  <word id="17" form="." lemma="punc1" postag="u-----" relation="AuxX" head="4"/>
  <word id="18" insertion_id="0001e" artificial="elliptic" relation="PRED" form="[adduxit]" lemma="adduco" postag="v3sria---" head="0"/>
  <word id="19" insertion_id="0012e" artificial="elliptic" relation="ATR" lemma="adduco" postag="v3sria---" form="[adduxit]" head="10"/>
</sentence>
```

Tüundra-haun näkymä (dependenssipuu ja taulukkonäkymä):

Visualization

Table view

token	lemma	Number	Case	Gender	Person	Tense	Mood	pos	xpos	deps	head	num	Speaker	VerbForm	Voice
Nos	nos	Plur	Acc	Masc				PRON	—		11	887	Agrius		
vero	verus	Sing	Abl	Neut				ADJ	—		1	888	Agrius		
,	,							PUNCT	—		4	890			
inquit	inquam	Sing			3	Pres	Ind	VERB	—		1	889	Agrius	Fin	Act
Agrius	Agrius	Sing	Nom	Masc				PROPN	—		4	891	Agrius		
,	,							PUNCT	—		7	895			
ut	ut							SCONJ	—		8	894	Agrius		
arbitror	arbitror	Sing			1	Pres	Ind	VERB	—		1	893	Agrius	Fin	Pass
,	,							PUNCT	—		10	897			
eadem	idem	Sing	Nom	Fem				PRON	—		11	896	Agrius		
causa	causa	Sing	Nom	Fem				NOUN	—		0	886	Agrius		
quae	quis	Plur	Nom	Fem				NOUN	—		13	899	Agrius		
te	tu	Sing	Abl	Masc				PRON	—		11	898	Agrius		
,	,							PUNCT	—		15	901			
rogatio	rogatio	Sing	Nom	Fem				NOUN	—		11	900	Agrius		
aeditumi	aeditumus	Sing	Gen	Masc				NOUN	—		15	902	Agrius		
.	.							PUNCT	—		4	892			

Show all attributes Save

Liite 2: Annotoinnissa käytetyt leimat ja relaatiot (UD-standardi)

Alkuperäiset ja täydelliset tiedot: <https://universaldependencies.org/guidelines.html>

Sanaluokkaleimat

ADJ	adjective
ADP	adposition
ADV	adverb
AUX	auxiliary
CCONJ	coordinating conjunction
DET	determiner
INTJ	interjection
NOUN	noun
NUM	numeral
PART	particle
PRON	pronoun
PROPN	proper noun
PUNCT	punctuation
SCONJ	subordinating conjunction
SYM	symbol
VERB	verb

Dependenssirelaatiot

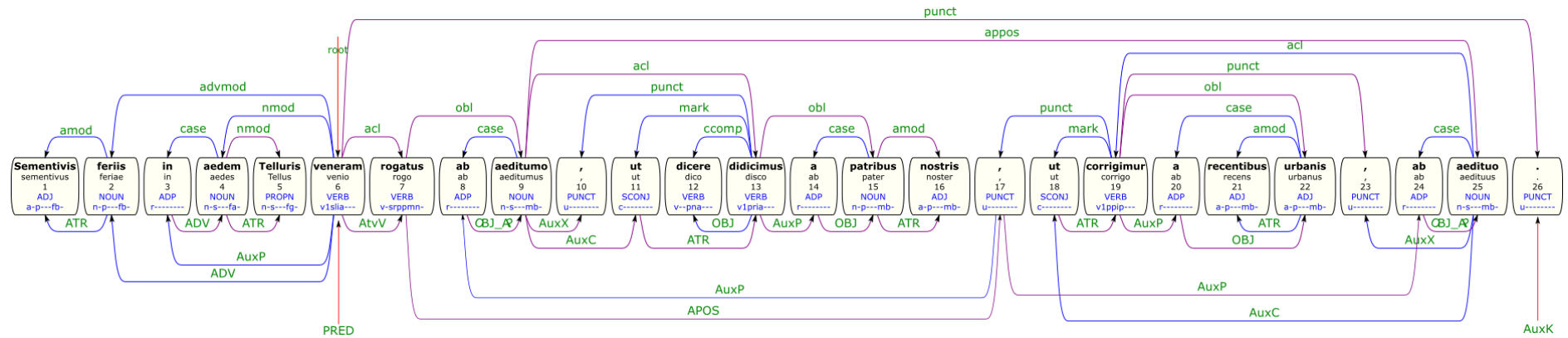
Dependenssirelaatioiden järjestelmä. Sulkeissa olevat relaatiot eivät ole käytössä tässä työssä käsiteltävissä puupankeissa.

	Nominit <i>Nominals</i>	Lauseet <i>Clauses</i>	Määritteet <i>Modifier words</i>	Funktiosanat <i>Function Words</i>
Pakolliset määreet <i>Core arguments</i>	nsubj obj iobj	csubj ccomp xcomp		
Vapaaehtoiset määreet <i>Non-core dependents</i>	obl (vocative) (expl) (dislocated)	advcl	advmod* discourse	aux cop mark
Nominimääreet <i>Nominal dependents</i>	nmod appos nummod	acl	amod	det (clf) case
Koordinaatio <i>Coordination</i>	Monisanaiset <i>MWE</i>	Löyhä <i>Loose</i>	Erikoistapaukset <i>Special</i>	Muut <i>Other</i>
conj cc	(fixed) flat (compound)	(list) parataxis	orphan (goeswith) (reparandum)	punct root (dep)

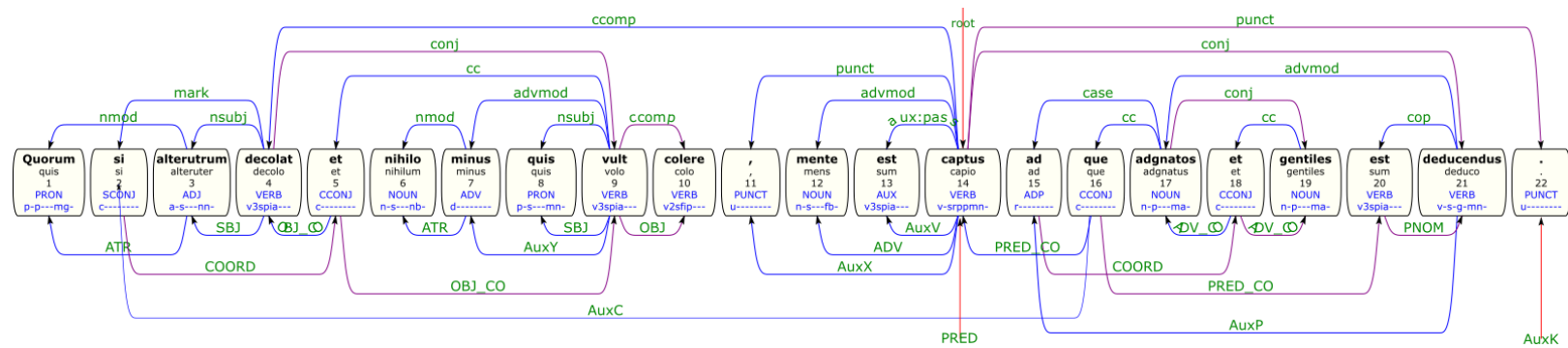
Relaatiolimat

acl	clausal modifier of noun (adjectival clause)
advcl	adverbial clause modifier
advmod	adverbial modifier
amod	adjectival modifier
appos	appositional modifier
aux	auxiliary
case	case marking
cc	coordinating conjunction
ccomp	clausal complement
clf	classifier
compound	compound
conj	conjunct
cop	copula
csubj	clausal subject
dep	unspecified dependency
det	determiner
discourse	discourse element
dislocated	dislocated elements
expl	expletive
fixed	fixed multiword expression
flat	flat multiword expression
goeswith	goes with
iobj	indirect object
list	list
mark	marker
nmod	nominal modifier
nsubj	nominal subject
nummod	numeric modifier
obj	object
obl	oblique nominal
orphan	orphan
parataxis	parataxis
punct	punctuation
reparandum	overridden disfluency
root	root
vocative	vocative
xcomp	open clausal complement

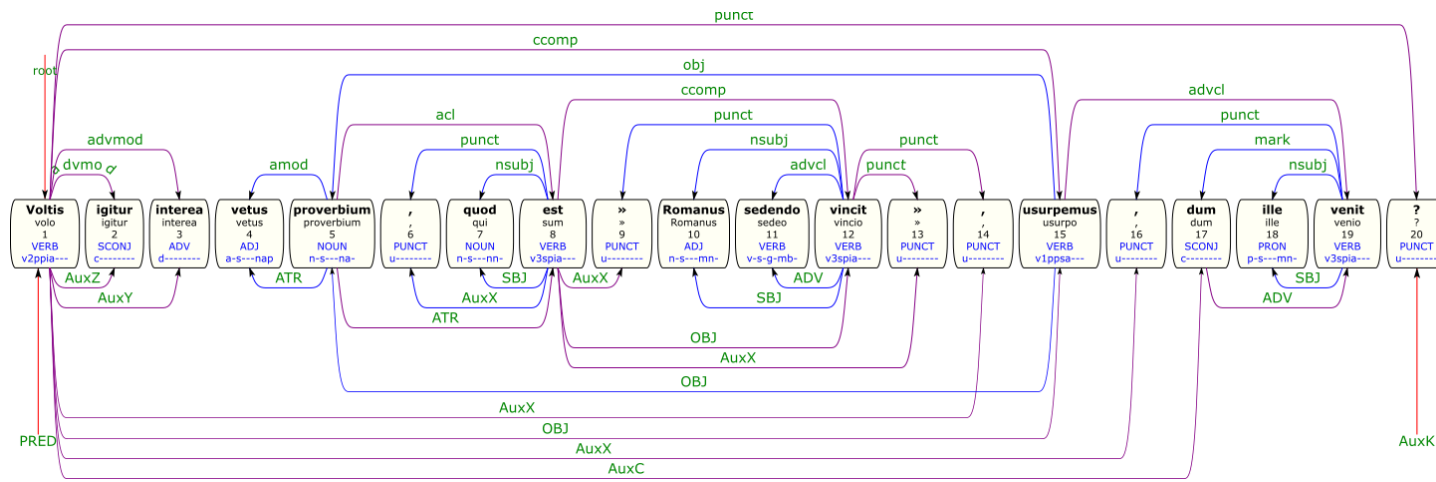
Liite 3: Annotoidut esimerkkivirkeet kokonaisina



Rust. 1,2,1



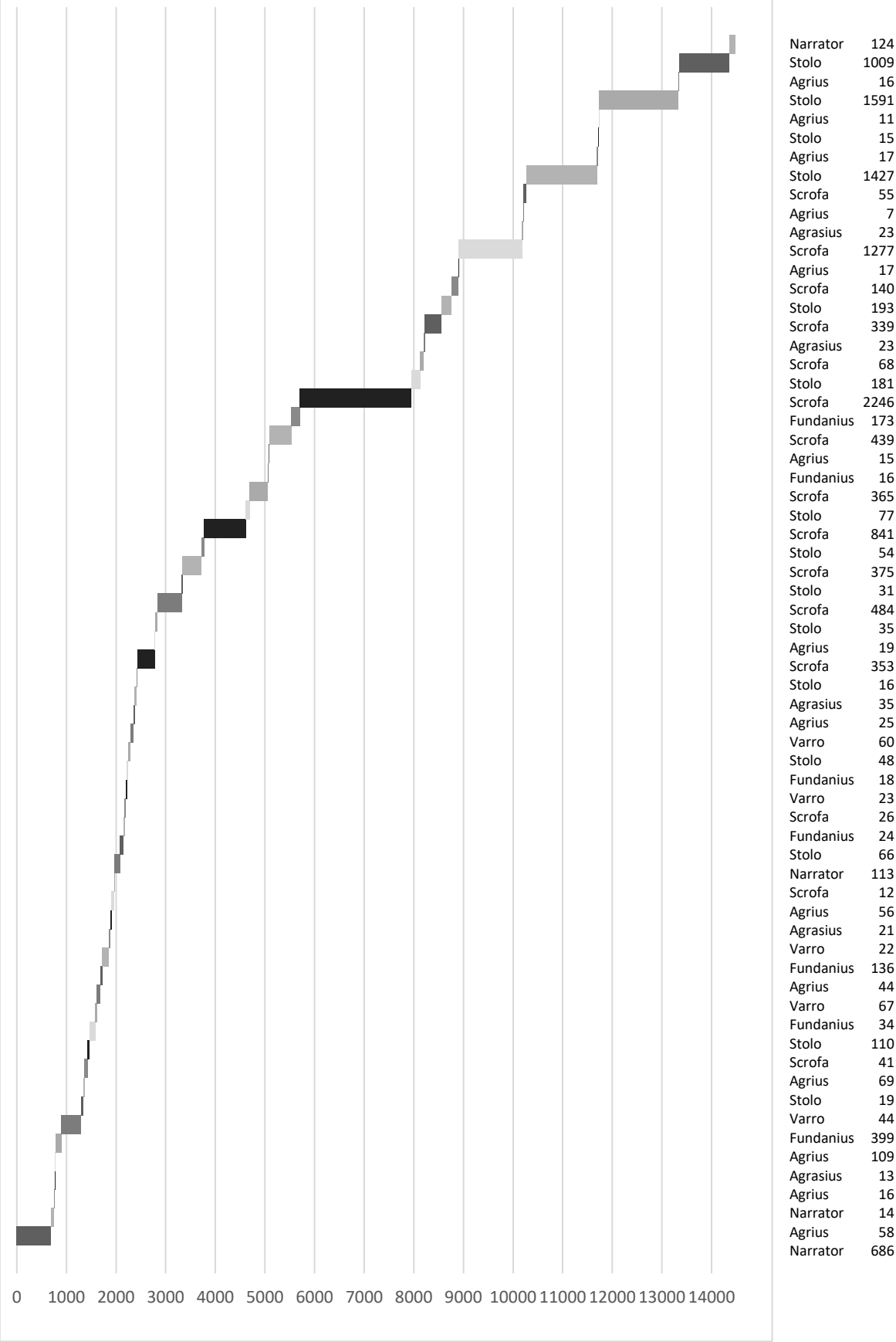
Rust. 1,2,8



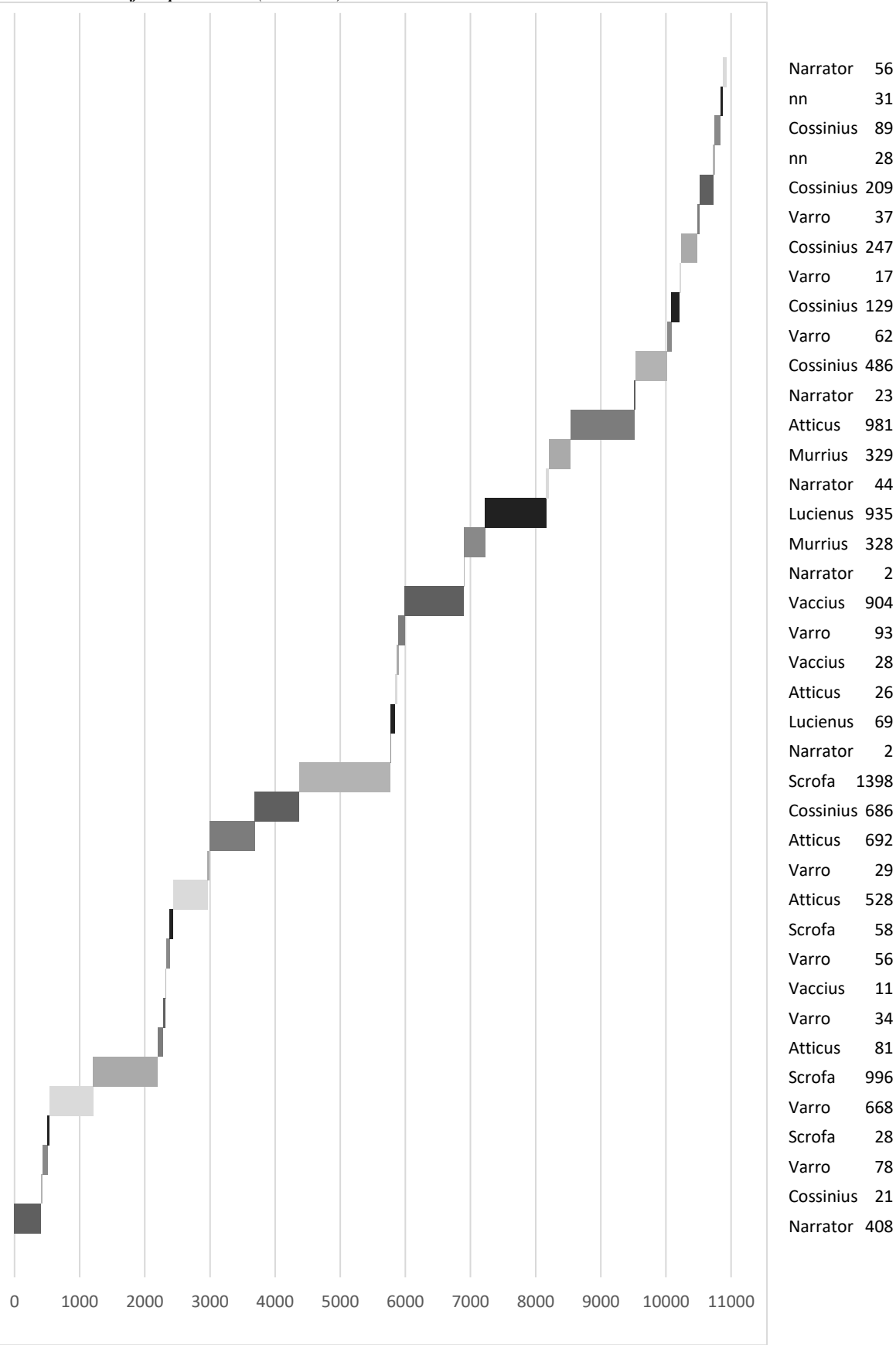
Rust. 1,2,2

Liite 4: Dialogien kulku: puheenvuorojen pituudet ja vaihtelu

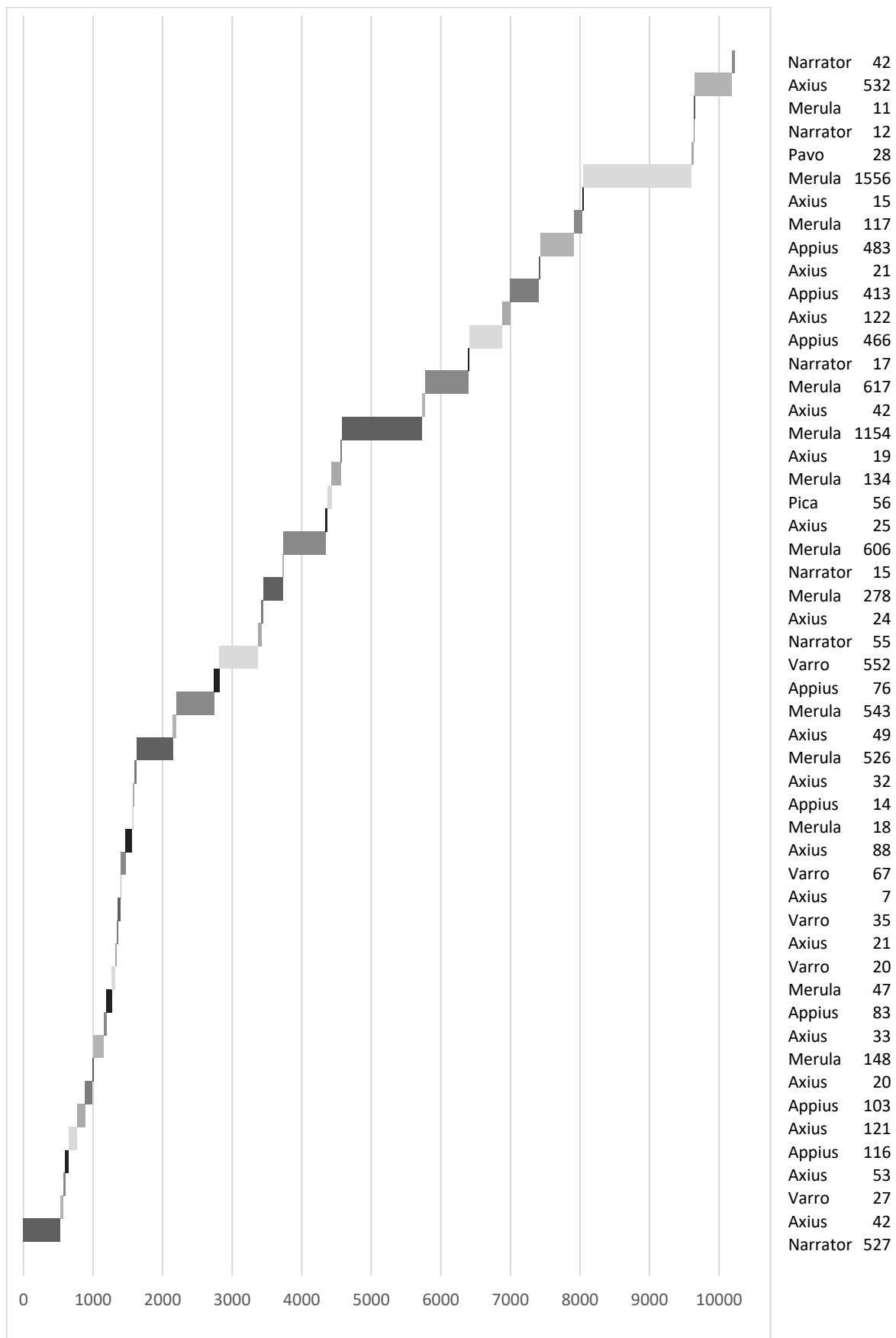
Puheenvuorojen pituudet (sanetta), Res Rusticae 1



Puheenvuorojen pituudet (sanetta), *Res Rusticae* 2



Puheenvuorojen pituudet (sanetta), Res Rusticae 3



Liite 5: Yleisimpien lemmojen esiintymät

lemma	määrä				prosenttiosuus			
	1. kirja	2. kirja	3. kirja	yhteensä	1. kirja	2. kirja	3. kirja	yhteensä
qui	557	363	352	1272	3,85	3,32	3,44	3,57
in	487	317	326	1130	3,37	2,90	3,19	3,17
sum	438	316	326	1080	3,03	2,89	3,19	3,03
et	411	365	278	1054	2,84	3,34	2,72	2,96
is	299	195	204	698	2,07	1,78	1,99	1,96
ut	291	196	177	664	2,01	1,79	1,73	1,86
ad	201	164	128	493	1,39	1,50	1,25	1,38
non	186	127	118	431	1,29	1,16	1,15	1,21
cum*	133	120	119	372	0,92	1,10	1,16	1,04
quod	159	115	89	363	1,10	1,05	0,87	1,02
hic	126	107	129	362	0,87	0,98	1,26	1,02
atque	153	96	110	359	1,06	0,88	1,08	1,01
ex	122	90	120	332	0,84	0,82	1,17	0,93
ab	97	146	74	317	0,67	1,34	0,72	0,89
ne	103	99	99	301	0,71	0,91	0,97	0,85
habeo	88	88	98	274	0,61	0,81	0,96	0,77
aut	120	55	82	257	0,83	0,50	0,80	0,72
de	98	99	54	251	0,68	0,91	0,53	0,70
si	118	75	56	249	0,82	0,69	0,55	0,70
que	92	78	77	247	0,64	0,71	0,75	0,69
alius	128	61	44	233	0,88	0,56	0,43	0,65
facio	91	66	66	223	0,63	0,60	0,65	0,63
enim	87	85	50	222	0,60	0,78	0,49	0,62
fio	88	67	51	206	0,61	0,61	0,50	0,58
dico	89	56	58	203	0,62	0,51	0,57	0,57
possum	87	58	52	197	0,60	0,53	0,51	0,55
quam	95	50	44	189	0,66	0,46	0,43	0,53
magnus	55	63	62	180	0,38	0,58	0,61	0,51
oportet	67	51	35	153	0,46	0,47	0,34	0,43
multus	51	44	39	134	0,35	0,40	0,38	0,38
etiam	59	42	32	133	0,41	0,38	0,31	0,37
ita	51	35	46	132	0,35	0,32	0,45	0,37
inquam	51	33	46	130	0,35	0,30	0,45	0,36
sed	57	35	36	128	0,39	0,32	0,35	0,36
locus	53	30	44	127	0,37	0,27	0,43	0,36
res	49	40	35	124	0,34	0,37	0,34	0,35
ubi	54	19	43	116	0,37	0,17	0,42	0,33
parvus	50	34	31	115	0,35	0,31	0,30	0,32
terra	85	10	12	107	0,59	0,09	0,12	0,30
soleo	32	38	32	102	0,22	0,35	0,31	0,29
pars	53	30	18	101	0,37	0,27	0,18	0,28
quis	42	28	31	101	0,29	0,26	0,30	0,28
ager	76	8	17	101	0,53	0,07	0,17	0,28
primus	42	30	28	100	0,29	0,27	0,27	0,28
genus	43	19	38	100	0,30	0,17	0,37	0,28